



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 112389

TO: Phuong Bui
Location: REM-2A15
Art Unit: 1638
Jan 22, 2004

Case Serial Number: 09831233

From: P. Sheppard
Location: Remsen Building
Phone: (571) 272-2529

sheppard@uspto.gov

Search Notes

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 19:35:55 ; Search time 7195 Seconds
(without alignments)
11121.505 Million cell updates/sec

Title: US-09-831-233A-1

Perfect score: 1956

Sequence: 1 gcacgaaggaactggca.....aaaaaaaaaaaaaaaaaaaa 1956

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_ey.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_ay.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	928.4	47.5	1996	8	AK118498 Arabidops
2	927.4	47.4	1779	8	BT005923 Arabidops
3	902.2	46.1	1994	8	AB006210 Arabidops
4	359.6	18.4	44308	8	AL033388 S.pombe c
5	336	17.2	2429	8	Z6117 S.cerevisia
6	303.2	15.5	87435	8	AF159463 Emericell
7	266.6	13.6	4351	8	Continuation (2 of
8	234.8	12.0	110000	2	AL402350 T3 end of
9	224.2	11.5	975	11	AF290177 Candida a
10	198.2	10.1	2306	8	AB016783 Arabidops
11	150.8	7.7	4581	8	AC134234 Oryza sat
12	129.6	6.6	195583	8	AL591975 Listeria
13	127.8	6.5	280050	1	AX641665 Sequence
14	127.8	6.5	349980	6	AX641666 Sequence
15	127.8	6.5	349980	6	AX641666 Sequence
16	127.8	6.5	349980	6	AX641666 Sequence
17	126.4	6.5	11440	1	AE014962 Streptoco
18	126.4	6.5	128294	8	AF377947 Oryza sat
19	126.4	6.5	137580	8	AL596165 Listeria
20	125.4	6.4	249050	1	AX417038 Sequence
21	125.4	6.4	349980	6	AX417041 Sequence
22	125.4	6.4	349980	6	AX417042 Sequence
23	117.2	6.0	4584	6	AX416785 Sequence
24	116	5.9	12075	1	AE013160 Thermoana
25	112.2	5.7	11308	1	AE007609 Clostridi
26	107	5.5	15451	1	AE000677 Aquifex a
27	106.2	5.4	128574	8	ATFL0M23
28	106.2	5.4	198372	8	ATCHRIV66
29	102.8	5.3	12027	1	AE000897 Methanoba
30	101.2	5.2	12634	1	AE000912 Methanoba
31	100.2	5.1	47739	1	AF017113 Bacillus
32	100.2	5.1	209510	1	BSUB0018
33	98.6	5.0	11141	1	U67493
34	98.6	5.0	110000	6	AR271569_10
35	98.6	5.0	110000	6	AR271569_11
36	96	4.9	8313	1	SSUS2227_11
37	96	4.9	13528	1	AE006689 Sulfolobu
38	96	4.9	281244	1	SSU18930
39	96	4.9	300150	1	AP004594 Oceanobac
40	95	4.9	10386	1	AE010714 Methanosa
41	94	4.8	306050	1	BX321858 Nitrosomo
42	93.6	4.8	10515	1	AE012824 Chlorobiu
43	92.4	4.7	10186	1	AE010316 Methanopy
44	92.4	4.7	3300	1	AF150930 Thermoana
45	92.2	4.7	3300	1	AF150930 Thermoana

ALIGNMENTS

RESULT 1

AK118498

LOCUS

DEFINITION

Arabidopsis thaliana At4g26900 mRNA for putative glutamine

amidotransferase/cyclase, complete cds, Clone: RAFL19-73-E07.

ACCESSION

AK118498

VERSION

AK118498.1 GI:26452023

KEYWORDS

FLI CDNA; CAP trapper.

SOURCE

Arabidopsis thaliana

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi

REFERENCE

1

Seki.M., Iida.K., Satou.M., Sakurai.T., Akiyama.K., Ishida.J.,

AK118498 1996 bp mRNA linear PLN 06-DEC-2002
Arabidopsis thaliana At4g26900 mRNA for putative glutamine
amidotransferase/cyclase, complete cds, Clone: RAFL19-73-E07.

Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayaehizaki, Y. and Shinozaki, K.
 Arabidopsis thaliana full-length cDNA
 Published Only in Database (2002)
 2 (bases 1 to 1996)
 Nakajima, M., Enju, A., Kamiya, A., Narusaka, M., Carninci, P., Kawai, J., Hayaehizaki, Y. and Shinozaki, K.
 Direct Submission
 Submitted (25-NOV-2002) Motoaki Seki, RIKEN Genomic Sciences Center; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: msek@gs.riken.go.jp, URL: http://pfigweb.gsc.riken.go.jp, Tel: 81-45-503-9625, Fax: 81-45-503-9586)
 An Arabidopsis full-length cDNA library was constructed essentially as reported previously (Seki et al. (1998) Plant J. 15:707-720; Seki et al. (2002) Science 296:141-145). cDNA cleaved with BamHI and XhoI was ligated to modified Lambda FLC-1-E vector (Carninci et al. (2001) Genomics 77:79-90) digested with BamHI and SalI. This clone is in a modified pBluescript vector.
 Please visit our web site (http://pfigweb.gsc.riken.go.jp/) for further details.

FEATURES

source

Location/Qualifiers

1..1996

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="4"

/clone="RAP19-73-E07"

/note="common name: thale cress"

1..1996

/gene="At4g26900/F10M23_240"

33..1811

/gene="At4g26900/F10M23_240"

/codon_start=1

/product="putative glutamine amidotransferase/cyclase"

/protein_id="BAC43102.1"

/db_xref="GI:26452024"

/translation="MEATAAPFSSIVSSRNQFSSSSIRASSPASLFLSKSGNVR
 KFKSPRLSVRSTSDSVTLDDYAGNVRISRNALRHGFISIKDVTGPDILNADR
 LIPGVGAFAPAMDVLRNTGMAELCKYIENDRPFLGICLGLQLLFDSSSENGPVKGL
 GVIPIGVNRPDASGIRVPHIGWALQVGDSEILDDVGNRVVFFVHSYRAIPSDENK
 DMSITCSGPFSSIRNRGNVAVQHPKESGEVGLSVLRRLHPKLPATOKEMEK
 ASKAKRVIACLDVRNDGLVQKQDIDVREQSNENVRNLKFPVLAQYIKDG
 ADEISFLNITGFRDPLQVLRQTKNVFPLTVGGIRDFDASGRYISLE
 VAAEYFRSGADKISIGSDVSAABEEFKSGVTKSLEQISRVYGNQAVVVIDPRR
 VVYNHPDDVPYKIVRTNPNPENGVEYARYQCTVSGSGREGRPIGAFELAKAVEELGAGE
 ILNLCIDCGQSGFDLDLKLISDSVGIPIVIASSGAGTDPHFSEVPEKTNASALAA
 GIFRKEVPISQVKEHLQESRIEVR"

555 a 386 c 491 g 564 t

BASE COUNT

ORIGIN

Query Match 47.5%; Score 928.4; DB 8; Length 1996;
 Best Local Similarity 73.8%; Pred. No. 2.3e-198;
 Matches 1210; Conservative 0; Mismatches 421; Indels 9; Gaps 2;
 QY 43 CAGAAATTCGCCGCCCTTGGCTGGTGCGAGCGGAGATTCTGTGTGATCTTTACTTTGATTA 102
 DB 179 CAGAAGCCTCTCCGCTCCGCGCATCTTCTACCTCAGATTCTGTGTGACTTGTGCTTGACTA 238
 QY 103 CGGTGCTGGAATCTTAGGAGTGTGGAGAACCCATCGCACACTTCGATTTGATATCAA 162
 DB 239 CGGAGCTGGAATGTTCGGAGATCCCGCAATGCTCTTCGTCTATCTCGGCTTCAGCATCAA 298
 QY 163 AGATGTGCAAAAGCAGAGGATATCTTAAATGCTAAGCGCCTTATCTTCTCGCGCTTGG 222
 DB 299 AGAGCTTCAACCGCGGAGACATCTCGAATGCTGATCGACTCATATTTCCAGCGCTGG 358
 QY 223 GGCCTTTGACCTCGAATGGATGTTCTTATTTGTAAGAGGCTGCTGAGACACTCTGTAC 282
 DB 359 TGCCTTTTGACCCGCCATGGATGTACTTAAACAGAACTGGATGGCTGAAGCTTGTGCAA 418

QY 283 TTACATTCAGATGATGACGACCTTTCTCTGGGTATATGCTGGGATGTCAGCTACTCTTTGA 342
 DB 419 ATATATTGAGATGACGCTCCATTTCTAGGCATATGCTTGGTCTACAACACTACTTTTGA 478
 QY 343 GTCAGTGAAGAAAATGGTCCAAATTCAGGCTCTGGCTTGGATTCCTTGAGCGGTTGGCG 402
 DB 479 TTTCTAGTGAAGAGATGGACCACTCAAGGCTCTGGTGTGATACCGGGAATAGTTGGACG 538
 QY 403 TTTTGAATCATCCAATGGTTTAAAGGTGACCATATTTGGATGCGATGCGCTTGGATATAA 462
 DB 539 CTTTGTATGCTTACGCTGGGTATTAAGAGTACCCCAATTCGGCTGGAAATGCTTTTGAAGTGG 598
 QY 463 GGAAGGCTCAGCAATTTTATGATGATGCTGGGGAATCAACATGCTGTTATTTTCTCACTCAT 522
 DB 599 GAAGGATTCGAATTTTGGATGATGCTGGAACCGCTCATGCTATTTTGTTCATTCGTA 658
 QY 523 TCAGGCAATATGCC---GAGGACAAAGAGTGGATTTTCATCTATCATGACGCTATGGTGA 579
 DB 659 CAGGGCAATTCATCATGATGAAATTAAGGACTGGATTTCTGCTACCTGTATTAATATGTTGA 718
 QY 580 CGATTTTATGTCATTCATTCAGAGGGAATGTTTCATGCACTGCAATTTTCATCCCGAGAA 639
 DB 719 ATCAITTTATATCTTCCATTAAGAGGGAATGTCATGCACTGCAATTTCCATCTCTGAAA 778
 QY 640 GAGTGGAGGTGTTGGACTTTTCCATATTTGAGAGATTTTTCGAATGCTGATTTCTTTTAAACA 699
 DB 779 GACGGGGAGGTGGGGCTTCTGTTTTAAGAGGTTCTTGCATCCAAAATTAACCT----- 833
 QY 700 CAAAGACAGAACCCCAATGAATCGAAAGGCTTCTAACTTTGCAAGAGAGATAATTTGCTTG 759
 DB 834 -GCAACACAGAGCCCAATGGAAGGAAAGGCTTCAAACTTTGCAAGAGGCTGATTTGCTTG 892
 QY 760 CTTTGTATGTCAGGGCAATGATTAATGGGATCTTGTGTTAAACCAAGGGAGCAATATGA 819
 DB 893 TCTTGTATGTCAGGACCAATGATTAAGAGATCTCGTAGTTTACTTAAAGGGGATCAGTATGA 952
 QY 820 TGTGAGAGACGTACAGAGAGATGAGGTGACAGAACCTTTGGCAAGCTTGTGAACCTTGC 879
 DB 953 TGTGAGAGACCAATCTAATGAAAACAGGTTCCAAACCTTTGGCAACCTTGTGATTTGGC 1012
 QY 880 TGGCGAGTATTTATTTAGACGGTCTGATAGGTCAGCTTCTTTAAACATTTACTGTTTCCG 939
 DB 1013 TGGCGAGTATTAACAAGATGGTCAGATGAGATTAGCTTTTAAACATACTATGATTTCCG 1072
 QY 940 GGACTTCCCTCTAGGCGATCTACCCATGCTACAGGCTCTTCAACGGCGATCTCGAAGCT 999
 DB 1073 CGATTTCTCTAGGGGATTTGCCGATGATTCAGGTTGTTGAGGCGACATCAAGAATATG 1132
 QY 1000 TTTTGTGCCATTAACCTGTCGGGGGTGGCATCAGGATTTTCTGATGCAAAATGGAAGGTA 1059
 DB 1133 CTTTGTACCACTTAACCTGTTGGAGTGGTATTAGAGACTTTACAGATGCTAGTGGCAGTA 1192
 QY 1060 TTATTTCTAGTCTAGAAATGGCTTCAGATTTTTCAGATCGGGCGCGCATTAAGGTTTCAT 1119
 DB 1193 CTATTTCTAGCTTGGAAATGCTGCTGAGTATTTTCAGATCCGGTCTGATTAAGATCTCCAT 1252
 QY 1120 CGGAGTGTAGCTTTTACACTCTCTGAGGAATATATTAACCCGGAGTGAAGCAGGAAA 1179
 DB 1253 AGGAAGTGAACGCTGTTTCTGCTGAGAGGATTCATAAAATCAGGGGTGAAGCAGGAAA 1312
 QY 1180 GAGCAGCATAGACAGATATCTACAGTATATGTTAAACCGAGCGAGTGGTTGTAAGCATTTGA 1239
 DB 1313 GAGTAGTTTGAACAGATATCCAGAGTTTATGGAATCAGGAGTGGTTGTTAGTATTTGA 1372
 QY 1240 TCTTCGCGAGTTTATCTTTGAGAAAACCCGATGAAGTAGAATTTTAAAGCCATCAAGTAAG 1299
 DB 1373 TCTTCGTAGAGTTTATGTGAAACCATCCGATGATGTGCCATACAAAGTCATCAGAGTAAC 1432
 QY 1300 CCATCCAGTCCAAACGGTGAAGAAATGCTGTTATGCTGCTGCTGCTGTTTAAATGCTGACG 1359
 DB 1433 TAAATCCAGGCCCAATGGAGAAATATGCCGGTATTCAGTGAACGCTGAGTGAGGACG 1492
 QY 1360 AGAAGGAGAGCCCATCCGAGCTTTATGAATAGCTTGTGAGGCTTGTGAGGAACCTTGGAGCTGG 1419

||||| 1493 AGAAGTCGACCTATTGAGCATTTGAGCTTGCAGAACGCGTTGAAGAAATTAGTGCCCG 1552
QY 1420 AGAAATATTATTGAATCGCATTTGATGATGTTCAAGGAAGAGTTCCGATATAGATCT 1479
Db 1553 TGAATATATTGAACTGATAGCTGATGTTGATGTTCAAGGAAGAGTTCCGATAGACTT 1612
QY 1480 GATCAAGCTAATATCCGATCTGTGAACTTCCTGTTATCCGACGACGCGTGCAGGAGT 1539
Db 1613 AGTAAAGCTCATCTCAGATTCAAGTACGATCGGATCGCAAGCAGTGCAGCAGGTAC 1672
QY 1540 CGCTGATCACTTCTCCGAGTCTTTAATGAACCAACGATCTGCTGCCCTTGCAGCTGG 1599
Db 1673 TCCGACCACTTTTCCGAGTGTGTTGAAGACCAACGATCTGCTGCCGCTTGTGTCGG 1732
QY 1600 CATTTTCCATCGCAAGAGGTTCCAAATTAAGGCTGTTAAAGAGCAGCTTGTGAAGGAGG 1659
Db 1733 CATTTTCCGCGGAAGAGGTACCAATCCATCTGTGAAGAGCAGTTCAAGAGGAGCG 1792
QY 1660 GATTGAAGTTAGATTGTAAG 1679
Db 1793 CATAGAAGTCAGGATCTGAG 1812

RESULT 2

BT005923 1779 bp mRNA linear PLN 18-MAR-2003
LOCUS Arabidopsis thaliana At4g26900 gene, complete cds.
DEFINITION BT005923
ACCESSION BT005923.1 GI:29028957
VERSION FLI CDNA.
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryote: Viridiplantae; Streptophyta; Embryophyta: Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 1779)
Chen, R., Chen, H., Kim, C.J., Shinn, P., Bowser, L., Carninci, P.,
Chan, M.M., Chang, C.H., Dale, J.M., Hayashizaki, Y., Huan, V.W.,
Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,
Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M.,
Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M.,
Southwick, A., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yamada, K.,
Yu, G., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A. and
Ecker, J.R.
Arabidopsis ORF clones
Unpublished

2 (bases 1 to 1779)
Chen, R., Chen, H., Kim, C.J., Shinn, P., Bowser, L., Carninci, P.,
Chan, M.M., Chang, C.H., Dale, J.M., Hayashizaki, Y., Huan, V.W.,
Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J.,
Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M.,
Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M.,
Southwick, A., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yamada, K.,
Yu, G., Yuan, S., Shinozaki, K., Davis, R.W., Theologis, A. and
Ecker, J.R.
Direct Submission
Submitted (18-MAR-2003) Salk Institute Genomic Analysis Laboratory
(SIGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN
Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGECC (SSP) Consortium members constructed and
sequenced the pUNI (ORF) clones using the RAFL cDNAs: Chen, R.,
Chen, H., Kim, C.J., Shinn, P., Bowser, L., Chan, M.M., Chang, C.M.,
Dale, J.M., Huan, V.W., Jones, T., Karlin-Neumann, G., Lam, B.,
Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J.,
Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,

Yamada, K., Yu, G., Yuan, S., Davis, R.W., Theologis, A., and
Ecker, J.R.

Chen, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)
contributed equally to this work as PIs.

FEATURES

source
1. .1779
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="4"
/clone="U60940"
/note="This clone is in pUNI 51
ecotype: Columbia"
1. .1779
/notes="putative glutamine amidotransferase/cyclase"
/codon_start=1
/product="At4g26900"
/protein_id="AA064858.1"
/db_xref="GI:29028958"
/translation="MEATAAPFSIVSSRQNFSSSSIRASSPASLFLSKSIGNVNR
KPKSPRSLSVRASSTSDSVVLLDYCAGNVRIRNALRHLGPSIKDVQTPGILNADR
LIPPGVAPAPAMDVLNRGTMAEALCKYIENDRPPLGICGLQLLSEENGSPVAGL
GVIPIGVRFDASAGIRVPHIGNALQVKDSEILDDVGNRHYFVHSYRAIPQPMGK
DWISCTNYGESFISIRRGNVHAVQHPKSEVGLSVLRFLPKLPATQPKMEKG
ASKLAKRVACLDVNDKDLVTKGDQYDVREOSNENENRNLGPKPLDLAGQYKDG
ADEISFLNTITGFRDPLGLPMIOVLROTSKNVFPLTVGGGIRDTDASGRYSSLE
VAAEYFRSGADKISIGSDAVSAAEFEIKSVTKSSLEQISRVYGNQAVVSDIPRR
VYVNHPPDVPYKIVRTNPNGBEETARYQCTVSGRGRPIGAPLAKAVBELGAGE
ILLNCIDCGQKGFIDLVKLISDSVGPVIASSGAGTPDHFSEVFEKTNASAAALAA
GIFHRKEVPIQSVKHEHLQERIEVRI"
BASE COUNT 487 a 347 c 453 g 492 t
ORIGIN
Query Match 47.4%; Score 927.4; DB 8; Length 1779;
Best Local Similarity 73.8%; Pred. No. 3.8e-198;
Matches 1209; Conservative 0; Mismatches 421; Indels 9; Gaps 2;
QY 43 CAGAAATTCGGCGCGCTTCGCTGGTCAGCGGAGATCTGCTGACCTTACTTCTTATTA 102
Db 147 CAGAAGCCTTCCTCGTCCGCGCATCTTCTACCTCAGATCTGTTGTGACCTTTGCTGAC 206
QY 103 CGGTGCTGGAATGTTAGGAGTGTAGAGAACGCCATCCGACACACTTGGATTTGATATCAA 162
Db 207 CGGAGCTGAATGTTTCGGAGCATCCGCAATGCTCTCTGCTCATCTCGGCTTCAGCATCAA 266
QY 163 AGATGTCAGAAACCCAGAGGATTTCTAAATGCTAAGCGCTTATCTTCTCGGCGGTGG 222
Db 267 AGACGTTCAAACGCCCGGAGACATTTCTGAATGCTGATCGACTCATATTTCCAGGCGTGG 326
QY 223 GGCCTTTGCACTGCAATGATGTTCTTATTCTGTAAGCGCTGCGCTGAAGCACTCTGTAC 282
Db 327 TGTCTTTTGACCGCGCATGGATGTACTTAAACAGAACTGGGATGGCTTTGTGCA 386
QY 283 TTACATTCAAGATGATCGACCTTTCTCTGGGTATATGCTGGGATTCAGCTACTCTTTGA 342
Db 387 ATATATTGAGATGACCGTCCATTTCTAGGCATATGCTTGTGCTACAACTACTTTTCGA 446
QY 343 GTCAAGTGAAGAAATATGTTCCAAATTCAGGCTTTGCTGCTTGAATTCCTGGAAGGCTTGGCG 402
Db 447 TTCTAGTGAAGAGATGGACCAAGTCAAAGGCTCTTGGTGTGATACCGGGAATAGTTGGACG 506
QY 403 TTTTGAATCATCAATGTTTAAAGGTGCCATATTTGATGGCATGCCCTTGGATATAAA 462
Db 507 CTTTGTATGCTTCAGCTGGTATAAGAGTACCCACATTTGGCTGGAATGCTTTGCAAGTTGG 566
QY 463 GGAAGGCTCAGCAATTTTATAGATGATGTTGGGAATCAACATGTGTATTTTGTTCACATCATA 522
Db 567 GAAGGATTTCTGAATTTTGGATGATGTTGGAACCCGTCAATGCTATTTTGTTCATTTCGTA 626
QY 523 TCGAGCCAAATGCC---GAGGACAAACAAAGAGTGGATTTTCATCTACATGCAAGCTATCGTGA 579

Db	627	CAGGGCCATTCCATCAGATGAAATAAAGGACTGGATTTCGCTACCTGTGAATTATGTGTGA	686
Qy	580	CGATTTATTTATGCAATCCATTCAGAAGGGAAATGTTTCATGCACTGCAAAATTTTCATCCCGAGAA	639
Db	687	ATCAATTTATATCTTCCATTAAGAGGGGAAATGTGCATGCAAGTTCAATTCATCTCTGAAA	746
Qy	640	GAGTGGAGGTGTGGACTTTTCCATPATTTGAGAGAATTTTTGAAATGCTGATTCCTTTAAACA	699
Db	747	GAGCGGGAGGTGGGGCTTTCGTGTTTAAAGAAAGGTTCTTGCAATCAAAATTAACCT-----	801
Qy	700	CAAAAGACAGAGGCAATGAATGAAAGGCTTCTAAACTTTGCAAGAGAGAGTAATTCGTTG	759
Db	802	-GCAACACAGAGGCAATGGAAGGAAAGGCGCTCAAAACTTTGCAAGAGGGTGAATTCGTTG	860
Qy	760	CTTTGATGTGAGGGCAATGATATATGGGATCTTGTGTTGAACCAAGGAGGACCAATATGA	819
Db	861	TCITTTGATGTGAGGACGAATGATAAGGAGATCTCGTAGTTACTAAAGGGATCAGTATGA	920
Qy	820	TGTGAGAGAACGTAACAGAGAGAATGAGGTACAGAAACCTTTGGCAAGCCTGTGTAACCTTGC	879
Db	921	TGTGAGAGAGCAATCTAATGAACAGAGGTTTCGAAACCTTTGGCAAACTGTGTTGATTTGGC	980
Qy	880	TGGCAGTAATTAATTAGCGGTGCTGATGAGGTCAGCTTCTTTAAACAATTAAGTTTCGCG	939
Db	981	TGGCAGTAATTAACAAGATGGTGAGATGAGATTAGCTTTTAAACAATAACTGGAATTCGCG	1040
Qy	940	GGACTTCCCTCTAGCGCATCTACCAATGCTACAGTCTGCAACCGCGCATCTGAAAACGT	999
Db	1041	CGATTTTTCCTCTAGGGATTTGCCGATGATTCAGGTGTGAGGCAGACATCAAGAATGT	1100
Qy	1000	TTTTTGTGCCATAACTGTCGGGGGTGGCATCAGGGATTTTACTGATGCAAAATGGAAGGTA	1059
Db	1101	CTTTGTACCATACTCTGTGGAGTGGTAATTAGAGACTTTACAGATGCTAGTGGCAGGTA	1160
Qy	1060	TTATTCCTAGTCTAGAAGTGGCTTCAGAGTATTTTCAGATTCGGGCGCCGATTAAGTTTCGAT	1119
Db	1161	CTATTCCTAGCTTGGAAAGTTGCTGCTGAGTATTTTCAGATTCGGTCTCTGATTAAGATCTCCAT	1220
Qy	1120	CGGAAGTGTAGCTTTTACATGCTGTGAGGATATATTTAAACCGGAGTGAAGACAGGAA	1179
Db	1221	AGGAAGTGAAGCTGTTTCTGCTGAGAGGAGTTCAATAAATCAGGGGTGAAGACAGGAA	1280
Qy	1180	GAGCAGCATAGACAGATATCTACAGTATATGGTAAACAGGCACTGGTTGTGAAGCATTTGA	1239
Db	1281	GAGTAGTTTAGAACAGATATCCAGAGTTTATGGAAATCAGGCAGTGGTTGTAGTATTTGA	1340
Qy	1240	TCCTCGCGAGTTTACTTTGAAAAACCCGATGAAGTAGAATTTTAAAGCCATCAAAAGTAAG	1299
Db	1341	TCCTCGTAGAGTTTATGTGAACCATCCGATGATGTCATACAAAGTCATCAGAGTAAC	1400
Qy	1300	CCATCCAGGTCCAAACGGTGAGGAATATGCTCGTATCAGTGCACTGTTAATGGTGCAGC	1359
Db	1401	TAATCCAGAGGCCAAATGGAGAAGAAATATGCCCGTATCAGTGCACGGTCTAGTGGAGGACG	1460
Qy	1360	AGAAGGAGACCCATCGAGGCTTATGAATAGCTAAGGCTGTTTCAGGAACTTTGAGCTGG	1419
Db	1461	AGAAGTGCACCTATTGGAGCAATTTGAGCTTTCGAAAGCGGTTGAGAAATTAAGTGC	1520
Qy	1420	AGAAATATTATGAACTGCATTTGATTTGATGGTCAAGGAAAGGATTCGATATAGATCT	1479
Db	1521	TGAAATACTATTGAATGCAATAGACTGTGATGGTCAAGGGGAAAGGATTCGACATATAGACTT	1580
Qy	1480	GATCAAGCTAATATCCGATGCTGTGAACATTCCTGTTATCGCAGCAGCGGTGCAGGAT	1539
Db	1581	AGTAAAGCTCATCTCAGATTTCAGTAGGCAATACCGGTGATTCGCAAGCAGTGGAGCAGTAC	1640
Qy	1540	CGCTGATCATCTTCTCCGAAGTCTTTTAAATGAAACCAACGCACTCTGCTGCCCTTCGAGCTGG	1599
Db	1641	TCCCGACCACTTTTCCGAGGTGTTTGAAGACAAACGCACTCTGCCGGGCTTGTGTCGG	1700
Qy	1600	CATTTTCCATCGAAAGGTTTCAATTTAAGGCTGTTTAAAGAGCACTTTGTTGAAGGAAGG	1659
Db	1701	CATTTTCCACCGGAAAGAGGTACCAATCCAACTCTGTGAAGAGAGCACTTACAAGAGGACG	1760

QY	1660	GATTGAAGTTAGATTGTA	1678
DB	1761	CATAGAAGTCAGGATCTGA	1779
RESULT 3			
AB006210			
LOCUS			
DEFINITION	1994 bp mRNA linear	PLN 05-FEB-1999	
	Arabidopsis thaliana mRNA for glutamine amidotransferase/cyclase,		
	complete cds.		
ACCESSION	AB006210		
VERSION	AB006210.1	GI:3219163	
KEYWORDS	glutamine amidotransferase/cyclase.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;		
	rosids; eurosoids II; Brassicales; Brassicaceae; Arabidopsiis.		
REFERENCE	1 (sites)		
AUTHORS	Fujimori, K. and Ohta, D.		
TITLE	An Arabidopsis cDNA encoding a bifunctional glutamine		
	amidotransferase/cyclase suppresses the histidine auxotrophy of a		
	Saccharomyces cerevisiae his7 mutant		
JOURNAL	FBS Lett. 428 (3), 229-234 (1998)		
MEDLINE	98316711		
PUBMED	9654139		
REFERENCE	2 (bases 1 to 1994)		
AUTHORS	Fujimori, K.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-AUG-1997) Ko Fujimori, Osaka Bioscience Institute,		
	Department of Molecular Behavioral Biology; 6-2-4 Furuedai, Suita,		
	Osaka 565-0874, Japan (E-mail:fujimori@obi.or.jp,		
	Tel:+81-6-872-4851, Fax:+81-6-872-2841)		
FEATURES	Location/Qualifiers		
source	1..1994		
	/organism="Arabidopsis thaliana"		
	/mol_type="mRNA"		
	/strain="Columbia"		
	/db_xref="taxon:3702"		
	6..1787		
	/codon_start=1		
	/product="glutamine amidotransferase/cyclase"		
	/protein_id="BA28783.1"		
	/db_xref="GI:3219164"		
	/translation="MEATRAFFSIVSSRQNFSSSIRASSPASLFLSQKSGNVR		
	KFSPRLSVRASSTSDSVTLDDYGAGNVRISRNALRHGFSIKDVQTPGDIINADR		
	LIPFGVPPAMPDLNRTGMALCKYIENDRPFLLICGLQLLPDSBQNGPVKGL		
	GVIPGIVGRDASGIRVPHIGNALQVKDSEILDDVGNHRVYFVHSVRAIPSDENK		
	DWISITVCSDESSITRRGNVHAVGFHPEKSGEVLVLRLFLHKLPLATQKPMGK		
	ASKLARVKIVLDRNTDKGLVTLTKGQDVREQNENRNLGKPDVLAGQYKDG		
	ADSLFINLTGFRDFLGDLPMTQVLRQTSKNVFLPTVGGGIRDFDTASGVYSLE		
	VAAEYFRSGAGKISDGAFAAEETFKSGVTGKSLQISRVYGNQA VVVISIDPR		
	VYVNHDPDVPKYIRVTNPGCEEYAWYQCTVSGGQGRPICAFELAKAVEBLAGE		
	ILLNCINCQGGGFDIDLKLLSDSGIPVIASSGAGTDPHSEVFEEDKRICRACC		
	RHPFFPERGYQSQSVKHLQERIEVRI"		
	1994		
	/note="17 a nucleotides"		
polyA_site	557 a 385 c 487 g 565 t		
BASE COUNT			

QY 163 AGATGTGCAAAAGCCAGAGGATATCTAAATGCTAAGGGCTTATCTTTCTGCGTTGG 222
DB 272 AGAGTTTCAAAAGCGCGGAGACATTTCTAAATGCTGATCGACTCATATTTTCCAGCGTTGG 331
QY 223 GGCCTTTGCACTGCAATGGATGTTCTTATTCGTAAGGGCTGCTGAAGCACTCTGTAC 282
DB 332 GCCTTTTGCACCCGCCATGGAATGATCTTAAACAGAACTGGGATGGCTGAAGCTTTGTCAA 391
QY 283 TTACATTTAGAAATGATCGACCTTTCTGCGGTATATGCTGCGGATTCGAGCTACTCTTTGA 342
DB 392 ATATATTAGAAATGACCGTCCATTTCTAGGCATATGCTTGGTCTACAACTACTTTTGA 451
QY 343 GTCAGTGAAGAAATGGTCAATTCAGAGTCTTTGGCTTGATCTCTCGACGGGTGGCG 402
DB 452 TTTAGTGAACAGAAATGGACAGTCAAGAGTCTTGGTGTGATACCGGGAATAGTTGGACG 511
QY 403 TTTTGAATCATCAATGTTTAAAGGTGACACATATTTGATGGCATGCCCTTGGATATAA 462
DB 512 CTITGATGCTTCAGCTGGTATAGAGTACCCACATTTGGCTGGGAATGCTTTGCAAGTTGG 571
QY 463 GGAAGGCTCAGCAATTTAGATGATGTGGGGAATCAACATGCTGATTTTGTCACTCAT 522
DB 572 GAAGATTTGGAATTTGGATGATTTGGAACCGTCACTGCTATTTTGTTCATTCGTA 631
QY 523 TCGAGCCAATGCC---GAGGACAAAGAGTGGATTTCACTACATGCAAGTATGCTGA 579
DB 632 CAGGGCCATTCATCAGATGAATAAGGACTGGATTTGCTCTACCTGTAATATGCTGA 691
QY 580 CGATTTTATGTCATTCATCAGAAGGGAATGTTTCATGCAAGTCCAATTTTATCCCGAGAA 639
DB 692 ATCATTTATATCTTCCATAGAGAGGGGGAATGTCATGCAAGTCCAATTTCCATCCTGAAA 751
QY 640 GAGTGGAGGTGTTGCACTTTCCATATTTGAGAAGATTTTGAATGCTGATTCCTTTTACAA 699
DB 752 GAGCGGGAGGTGGGGCTTTCTGTTTAAAGAGTCTTGCATCCAAAATTAACCT----- 806
QY 700 CAAAGACAGAGCAATGAATGAAGAGGCTTTTAAACTTTCGAAAGAGATTAATGCTTG 759
DB 807 -GCAACAGAGAGCAATGGAAGGAGGAGGCTTCAAACTTTCGAAAGAGGATGATGCTTG 865
QY 760 CTTGATGTAGGCAATGATATGAGGATCTTGTGTAACCAAGGAGACCAATATGA 819
DB 866 TCTTGATGTAGGAGCAATGATAAGGAGATCTGTAATGTAATTAAGGGGATCAGTATGA 925
QY 820 TGTGAGAGACGTACAGAGAGATGAGTGCAGAACTTGGCAAGCTTGTGAACCTTGC 879
DB 926 TGTGAGAGAGCAATCTAATGAACAGAGTTCGAAACCTTGGCAACCTGTTGATTTGGC 985
QY 880 TGGGCACTATTTAGACGCTGCTGATGAGGTGAGCTTCTTAAACATTAAGTCTTCCG 939
DB 986 TGGGCACTATTTACAAAGATGCTGCAGATGAGATTTAGCTTTTAAACATTAAGTCTTCCG 1045
QY 940 GGACTTCCCTTAGGCGATCTACCCATGCTACAGGTCTTGAACCGCGATCTGAAACGT 999
DB 1046 CGATTTTCTCTAGGGATTTGCGGATGATTCAGGTGTTGAGGAGACATCAAGAAATGT 1105
QY 1000 TTTTGTCCATTAATCTGCGGGGTGCAATCAGGATTTTACTGATCAATGGAAGGTA 1059
DB 1106 CTTTGTACCACTAACTGAGGTGATATAGAGACTTTTACAGATGCTAGTGGCAGGTA 1165
QY 1060 TTATTTCTAGTGTAGAGTGGCTTACAGATATTTTACAGATCGGGCGCCGATTAAGGTTTCAT 1119
DB 1166 CTATTTCTAGTGTGAGGTGCTGCTGATTTTACAGATCCGGTCTGATTAAGATGTCAT 1225
QY 1120 CGGAAGTGTAGTGTACCTGCTGAGGAAATATATTTAAACCGGAGTGAAGACAGGAAA 1179
DB 1226 AGGAAGTGAAGCTGCTTTTGTGTCAGAGGAGTTCATTAATTCAGGGGTGAAGACAGGAAA 1285
QY 1180 GAGCAGCATAGAGCATATCTACAGTATATGTAACCGGAGTGGTCTTAAGCATTTGA 1239
DB 1286 GAGTAGTTTGAACAGATATCCAGGTTTATGGAATTCAGGCAAGTGGTGTGAAGTATTTGA 1345
QY 1240 TCCTCGCCGAGTTTACTTTGAGAAAACCCGATGAAGTAGAATTTTAAAGCCATCAAGTAAG 1299

DB 1346 TCCTCTAGAGTTTATGTGAACCATCCGATGATGTGCCATACAAAAGTCATCAGAGTAAC 1405
QY 1300 CCATCCAGGTCCAAACCGGTGAGGAATATGCTGATATCAGTGACACTGTTTAAATGTTGAGC 1359
DB 1406 TAATCCAGGCCCAATGGAAGAAATATGCTGATATCAGTGACCGTCACTGAGGAGCA 1465
QY 1360 AGAAGGAGACCATCCGAGCTTTATGAATAGTAAGGCTGTTGAGAACTTCGAGCTGG 1419
DB 1466 AGAAGGTTCGACCTATTGGAGCATTTGAGCTTGCAGAAAGCGTTTGAAGAAATTAGGTGCCG 1525
QY 1420 AGAATATTATTGAATCGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1479
DB 1526 TGAATATCTATTGAATCGATTAATCTGATGATGATGATGATGATGATGATGATGATGATGAT 1585
QY 1480 GATCAAGCTAATATCCGATGCTGTAACATTCCTGTTATCGCAAGCAGCGTGCAGAGT 1539
DB 1586 AGTAAGCTCACTCAGATTCAGTAGCATACCGGTGATCGCAAGCAGTGGACAGGTAC 1645
QY 1540 CGCTGATCACTTCTCCGAAAGTCTTTAATGAA-ACCAACGATCTGCTGCCCTTGCAGCTG 1598
DB 1646 TCCCGACCACTTTCCGAGGTGTTTGAAGAGACAAACGATCTGCGCGCTTGTGCTGCG 1705
QY 1599 GCATTTTCCATCCAAAGAGGTTTCCAATT--AAGGCTGTTTAAAGAGCACTTGTGAAGGA 1656
DB 1706 GCATTTTCCACCGAAAGAGGTTACCAATCCCAATCTGTGAAGAGCACTTACAAGAGGA 1765
QY 1657 AGGGATTGAAGTTAGATTGTAAG 1679
DB 1766 GGCATAGAAGTCAGGATCTGAG 1788

RESULT 4

SPBC887/c
LOCUS SPBC887 44308 bp DNA linear PLN 01-JUL-1999
DEFINITION S.pombe chromosome II cosmid c887.
ACCESSION AL033388

VERSION 3-oxoacyl-[acyl-carrier-protein]-synthase; 50s ribosomal protein
114; adenylate cyclase; amidotransferase; calcium-transporting
leucine channel; golgi membrane; histidine biosynthesis;
pepase Rich repeat; lipid metabolism; mcs4; nuclear division;
pif1; protein sorting; pseudouridylylase synthase; response regulator
mcs4; rna-binding; rpl1; rrm3-pif1 helicase homolog;
transcriptional activator; WD repeat.
SOURCE Schizosaccharomyces pombe (fission yeast)
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
1 (bases 1 to 44308)

REFERENCE

1 (bases 1 to 44308)
Lynne, M., Rajandream, M.A., Barrell, B.G., Lelaure, V. and Galibert, F.
Direct Submission
Submitted (04-NOV-1998) European Schizosaccharomycetes genome
sequencing project, Sanger Centre, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
and Laboratoire de Recombinaisons Genetiques, 2 avenue du
Professeur Leon Bernard, Faculte de Medecine, F-35043 Rennes cedex,
France

COMMENT

Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL: http://www.sanger.ac.uk/projects/s_pombe/)
During 1995 to 1996 about 66% of S. pombe chromosome 1 was
sequenced by the Sanger Centre. The sequencing of the S. pombe
genome is now being continued with funding from The European
Commission. Fourteen European sequencing laboratories, including
the Sanger Centre, are participating in the project.
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in FomBase (an ACSDS
database) with additional predictions for the branch-acceptor sites
supplied by the program Sp3splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the

number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.
 CDS are numbered using the following system eg SPBC5H2.01c. SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c (complementary strand).
 The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
 The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
 Cosmid c887 has no overlaps at present.

FEATURES

source

1. 44308
 /organism="Schizosaccharomyces pombe"
 /mol_type="genomic DNA"
 /strain="972h-"
 /db_xref="taxon:4896"
 /chromosome="II"
 /map="IIR"
 /clone="cosmid c887"
 790..1326
 /gene="SPBC887.01"
 790..1326

gene

CDS

/gene="SPBC887.01"
 /note="SPBC887.01, len:178, SIMILARITY:Arabidopsis thaliana, O2327, hypothetical 108.0 kd protein, (940 aa), fasta scores: opt: 350, E():5.6e-27, (57.1% identity in 84 aa)"
 /codon_start=1
 /label="SPBC887.01"
 /product="hypothetical protein"
 /protein_id="CAA21886.1"
 /db_xref="GI:3850097"
 /db_xref="SPTREMBL:O94286"
 /translation="MRAVIFQDEGDKPNDSEAELEAAKVSRYRHHGDLHTF ADGLMKYGRNDEVVSKGLGDRYDNNVKKPFEEHLHEDEIRLLDNGYFVR ELDARKSYLNSIRKSVVY"
 join(1830..1959,2071..3944)
 /gene="SPBC887.02"
 join(1830..1959,2071..3944)
 /gene="SPBC887.02"
 /note="SPBC887.02, len:667, SIMILARITY:Cavia porcellus, P97 274, volume-regulated outwardly-rectifying chloride channel, (760 aa), fasta scores: opt: 450, E():0, (26.8% identity in 724 aa)"
 /codon_start=1
 /label="SPBC887.02"
 /product="putative chloride channel protein"
 /protein_id="CAA21887.1"
 /db_xref="GI:3850098"
 /db_xref="SPTREMBL:O94287"
 /translation="WKDDLLIDWHERYEQKSNRAGRFGLGLETKYSIISIG IFGLTALLNALSLNSREGCTWNILPDQCTCTEDYCECFEPMNNHVS FVCLIIYVSQVGFATTLGYVAPARASQPTTINAILSGKYFDMNVFSIKTL GSAELVCFVSAGLWKGKGFVHIATNIYILVERIAPSLADESEIFRQLAAWAS G1AASFAPGVGVIFFALEQLASSPFLFTGSIWYEFCLSSAVVALQLIRSWHTDVG IYVSVIDRWYSYKOTLPFIPISILCGLAGSLIYLNKPKASKTGFSKISNVFFVIF LSLTSLTADWYCESELLFNPMLFPQVINCSPSSSTVLCTTFWVAIVLFTSAL LGLLTATGGAAPPTGIIIVPSLAIGACIGRAVGTGLKSPPLSAGTSYGVIGSIAP LSSSTRIVVALVILFELTGAINTALPMLATLISKWVSIIETSIYDAWIKQPNIF YFPSSSLKFSIPLNFRVSPQLVRPIRSCSIEELERAMHDSQSFFVVLKNDTFE PFGFISRNKYSLELNRPMSNMOTDNTGLDPLRSAPFVDSFTDLFDYIHPTFTFL NYDTPPVLMKLKFDAGITNLALLNHGKLHGLVTKIVCFICTI"

misc_feature

/gene="SPBC887.02"
 /note="Pfam match to entry PF00654 voltage_CLC, Voltage gated chloride channels"
 1960..1965

misc_feature

/gene="SPBC887.02"
 /note="gtacgt, splice donor sequence"

misc_feature

gene

CDS

2055..2070
 /gene="SPBC887.02"
 /note="ctgacattttggaag, splice branch and acceptor"
 complement(join(4238..6370,6416..6526))
 /gene="SPBC887.03c"
 complement(join(4238..6370,6416..6526))
 /gene="SPBC887.03c"
 /note="SPBC887.03c, len:747, SIMILARITY:Saccharomyces cerevisiae, O07896, chromosome xii reading frame orf ylr002c., (663 aa), fasta scores: opt: 1099, E():1.9e-32, (33.1% identity in 629 aa)"
 /codon_start=1
 /label="SPBC887.03c"
 /product="hypothetical protein"
 /protein_id="CAA21888.1"
 /db_xref="GI:3850099"
 /db_xref="SPTREMBL:O94288"
 /translation="MAARKNQSPKPKVASSKLNKIKSKNNSSOSTEPKRNATSL DSKTKTKGVALPEIAARELTQEDIEFFNPFSSKYLSSINPDLGKVEKGRPDID YLKKSQQFELDTSLRSSDESVLDYKDSQEDQYELRPRVSSWNNSYNRLPIKT KQGLQNVADVNGEEFLESESEASLEIDSIDKEKOKSLEEQKIAPFIPVKQOIK NDKALGIAOQQLPEEPVENLHIRNIPKFPSPYTIKKLSILLTLLAVPRDIIKGYK IRPLSEEGTKLSKEVAORWEYEQTLKHYAKFLQTLLETILKSPSSTIDETQLSLYQ VAVRCCTKLEQASHENLSEKLFALAVRQISHTKPGDGIINSUKNIFBEDNGLK SLKCVTILSRMFKQRYNDVLDVDFLSVNLNDMKIDEAWQDDTTTFKRRKOLP YLTARKAYKETKITQEMKEADAVITQADKEQYSEILKIFITFYFTLQKGLKI GNALEGAARLSHLNTEFLGDLLOVLRELVDKTVFLPKDQSGVQATREALLTSTAF ETASAGVQKLNLDLGLFVORLYKLIIFPFLNPADNLNLTKRLKDPADSPKPVV NATTEMMLKCFQVFFKSNISRSLSFKRLAISMQLPHEHSAADLALLKKLL SYSKLSRLLTSEEQIGDGIYNPFIEDPLSNSSTAVLYEPFLKKNHSPAVSQSAKE LKKTSL"

misc_feature

misc_feature

gene

CDS

complement(6371..6381)
 /gene="SPBC887.03c"
 /note="cttacatttag, splice branch and acceptor"
 complement(6410..6415)
 /gene="SPBC887.03c"
 /note="gtatga, splice donor sequence"
 complement(join(6806..8727,8788..8959,9026..9073))
 /gene="SPBC887.04c"
 complement(join(6806..8727,8788..8959,9026..9073))
 /gene="SPBC887.04c"
 /note="SPBC887.04c, len:713, SIMILARITY:Mus musculus, PLAP MOUSE, phospholipase a-2-activating protein, (646 aa), _fast a scores: opt: 1091, E():0, (41.8% identity in 402 aa)"
 /codon_start=1
 /label="SPBC887.04c"
 /product="WD repeat-containing protein"
 /protein_id="CAA21889.1"
 /db_xref="GI:3850100"
 /db_xref="SPTREMBL:O94289"
 /translation="MTSYELSRBLGCHKQDVRGVCISNELIGSASRDGTVSYWEQIN GEWTFPHNHEGFVNCVCPVPAIDKNGRGQKCGILQEVGTNRSYLYFGHESNIC SASALNETIITGSDNSTARVMALGQCKYVLGQSSVAVLALGEDIFITGSAKLI KWNGKLVKSLIAHNDVRCQIPGFGASCNDGVIKLTSDGDFLHGHGTSFV YSLYTHNQILASCCEDVRTIRWKGECLOCITLPTTSVWSVSLPGLVHGTGSDG FVPIFTVDRVAPTEVLNFERVSOFAISSQVGIKKGLSLPGLIILSKPGKADQD VNVRVNDVEAYQWQKNEKKICQVDAVGNRKLPEGEKEYDVFVDVADQQA PLKPLNATENPYQAANRELNLQPLSYDVEVVKTEKTHQHSLESKEPNLESQS PLKPKITTFVSQLDSFNANVPAMCORLRSNLTNKNPLPAKSIDSLERALSCKITD TENKELLETCLSLDSLAERFPALDARLLAINSSDLAPIFLEVFVRVKSVPSS GSINVMVLAHGLSNVVPNTDAEGVSKMDCLSTVPQASSAKDFKIAFATLANN LSTLLQLALENTGIELLSILFSLDPPDNEAFYRALMALGTLCTVPDIALAASQI YHAQSVHGAERFSQEMFVDAEKILSLF"

misc_feature

complement(8228..8341)
 /gene="SPBC887.04c"
 /note="Pfam match to entry PF00400 G-beta, WD domain, G-bet a repeat"

misc_feature

complement(8345..8461)
 /gene="SPBC887.04c"
 /note="Pfam match to entry PF00400 G-beta, WD domain, G-bet a repeat"

misc_feature

complement(8468..8578)

```

/misc_feature
/genes="SPBC887.04c"
/notes="pfam match to entry PF00400 G-beta, WD domain,
G-bet a repeat"
complement(8582..8695)
/genes="SPBC887.04c"
/notes="pfam match to entry PF00400 G-beta, WD domain,
G-bet a repeat"
complement(8699..8812)
/genes="SPBC887.04c"
/notes="pfam match to entry PF00400 G-beta, WD domain,
G-bet a repeat"
complement(8728..8751)
/genes="SPBC887.04c"
/notes="tgcacaaagggttacatatatttcag, splice branch and
acceptor"

Query Match      18.48; Score 359.6; DB 8; Length 44308;
Best Local Similarity 55.1%; Pred. No. 2.8e-70;
Matches 848; Conservative 0; Mismatches 659; Indels 33; Gaps 6;

QY 177 CAGAGGATATCTAAATGCTAAGCGCTTATCTTCTCGGCTTGGGGCTTTGCACTG 236
DB 44191 CAATACTAATCTATCTTCACTGCTGATTTTCTCGTGTGGAACTTTGGTTTCG 44132

QY 237 CAATGGATGTTCTTATTCGTAAGGCTGCTGAAGCACTCTGTACTTACATTCAGAATG 296
DB 44131 TATGCGATTCTTAGCAAGCAAGGATTTTGAACCCCTTGGCTGCTATGCTTTGTCG 44072

QY 297 ATGACCTTTCTGGGTATATGCTGGATGCGACTACTCTTTGAGTCAAGTGAAGAAA 356
DB 44071 GCAACCTTTTATGGCTGTATGCTGTGAATCCAAGCATTAATTCGAGGGGCTGTGTAAG 44012

QY 357 ATGCTCAATCTCAAGGCTTGGCTTGAATCTCTGACGGTGTGGCGTTTGAATCATCCA 416
DB 44011 CACCACATCTAAAGGCTGGCGTATTTCCGGGTAGTCCAAAGGTTGA---CAAGC 43955

QY 417 ATGCTTTAAGGGTGCCCATATTTGGATGGCATGCTTTGGATATATAAGGAAGGGTGACAA 476
DB 43954 ATGACAGACCGTCCCTCATATTTGGTTGGAATCTCTGTGCTGATGACACTTCA 43895

QY 477 -----TTTGTAGATGATGGGNAATCAACATGTGATATTTTGTTCACATATTCGAGCCA 530
DB 43894 AAGAGTTTTTTTGTGTATGCGCCCATGACAAAATTTATTTTGTGCACTTTATATGATTC 43835

QY 531 ATCCGAGGACAA-----AGAGTGGATTTTCATCTACATCAGCTATGG---TG 578
DB 43834 CAGAAAAGGAGCTTATCTCTCCAGAGTTCAAAATTTGCTACCAAGAAATATGGAATG 43775

QY 579 ACGATTTTATGTCATTCATTCAGAGGGAATGTTCAATGCACTCCAAATTTTCATCCCGAGA 638
DB 43774 AAACCTTTGTAGGTGCTATTTGTAAGAAATAATTTTGTAGTACCAATTTTCATCCGGAAA 43715

QY 639 AGAGTGAGGTGTTGGATTTCCATATGAGAGATTTTGTGAATGCTGATTCCTTTAACA 698
DB 43714 AAAGTGTATCTGAGGTTTGGTTGTTTAAAGGCAATTTTAAACCGGTAATATGAGCAAC 43655

QY 699 ACAAGACAGAGACCAATGATGGAAGGCTCTTAACTTG-----CAAGAGAG 749
DB 43654 CTATAGTGTGAGGATCTAAACTTATAGAGAACTCTTTTGGCGGATTAACAAAGGTA 43595

QY 750 TAATTGCTGCTTGTATGAGGGAATATGATTAATGGGATCTTTGTTGAACCAAGGAG 809
DB 43594 TTATTGCTTGTAGATGTCGTTTCCAAACGACGCTGTGACCTTTGTAGTTTACAAAGGTG 43535

QY 810 ACCAATATGATGTAGAGAAACGTTACAGAGAGATAGGTCAGAAACCTTTGGCAAGCTG 869
DB 43534 ATCAATATGATGTCGGAATAATCTTCTGGAAGCGAGGTAGAAATCTTTGGAATAACCTG 43475

QY 870 TTGAACTTGTGGCAGTATTTTACACGGTGTGATGAGTCAAGTCTTTTAAACATTA 929
DB 43474 TGGAGTTATGTCAAAAGATATTTTCAAGAGGCGCTGACGAAGTCTGTTTTTGAACATCA 43415

QY 930 CTGTTTCCGGGACTTCCCTCTAGGCGATCTACCCATGCTACAGGTCTTTCGAACGCGCAT 989

```

```

DB 43414 CTTTCAATTAGAATTTGTCCAATGGCGGATGCACCTATGTTTACAGGTTTTAGAAAAAGCGAG 43355
QY 990 CTGAAAACGTTTTTGTGGCCATTAACTGTGCGGGGTGCGCATCAGGATTTTACTGTATGCA 1049
DB 43354 CTCAACAGATTTTGTACCTTTTAAACAGTAGTGTGGTGAATCAGAGATGTATCAGATCCTG 43295
QY 1050 ATGGAAGGTATTAATCTTAGTCTAGAAAGTGGCTTTTCAAGATATTTTTCAGATCGGGCCG 1109
DB 43294 ATGGGACTTTTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43235
QY 1110 AGTTTTCGATCGGAAGTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1169
DB 43234 AAGTATCTATGGAAGTGTGCGGCTTTATGCGCGAGAAAGTATTAACGAAATTTGGGAAGA 43175
QY 1170 A---GACAGAGAAAGACAGCATAGACAGATATCTACAGTATATGTTAAACCGCAGCTGG 1226
DB 43174 AATTGAGTGAAATAACTGCTATAGAAACAAATTTCTAAGGCTATGTTAATCAGCAGTTG 43115
QY 1227 TTGTAAGCATTTGATCTCGCCGAGTTTACTTTGAGAAAAACCCGATGAAGTAGAATTTAAAG 1286
DB 43114 TTATTTTCGTTGATCTTAACCGTCACTATGTAAGTTCCTGAAGATCTTAACACCATG 43055
QY 1287 CCATCAAGGTAAAGCCATTCAGGTTCCAAAGGTGAGGAATATGCTGCTGCTGCTGCTGCTG 1346
DB 43054 TTGTGAAAAACATCTCGGCTTGGACCTTAATGGAAGACATATTTGTTGCTATCAGTCACCG 42995
QY 1347 TTAATCGTGGAGAGGAGGAGACCCATCGGACTTTTGAACCTATGAACCTAGTAAAGCTTTGAGG 1406
DB 42994 TTAAGGAGTCTGTAATATCGGTATGATGTTGTTGAACCTTACTCGGCTTGTGAAG 42935
QY 1407 AACTCGGAGCTGGAGAAATATTTATGAACCTGCAATGATTTGTTGTTGTTGTTGTTGTTGTTG 1466
DB 42934 CAATGGGTCTGTTGAGTACTATTAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 42875
QY 1467 TCATATAGATCTGATCAAGCTTAATATCCGATGCTGTTGAACATTTCTGTTTATCGCAAGCA 1526
DB 42874 ATGACATAGAACTTTGTTGCTTTGTTTAAAAAATCTGCTGAATATTTCCAGTAATAGCTTCAA 42815
QY 1527 GCGGTCGAGGCTGCTGATCACTTCTCGGAGTCTTTTAAAGGCTTTTAAAGGCTGCTGCTG 1586
DB 42814 GTGGCTGCTGTTATTTCCCAACATTTTGAAGAGTATTTTAAAGAAACGGATTTGTGTGTCAG 42755
QY 1587 CCCTTCGAGCTGGCACTTTTCCATCGCAAGAGGTTTCCAAATTAAGGCTTTTAAAGAGCACT 1646
DB 42754 CTCTTCGCGCGGAATTTTTCATCGCAAACTTTGTCGATAGAGAGCTCAAGAGATATC 42695
QY 1647 TGTGTAAGAGGAGTGAAGTTAGATTGTAAAGCGAGAA 1686
DB 42694 TGGCAATACAGCAGCTTCTGTTGCGCACTTAGCTAGCGAA 42655

```

```

RESULT 5
SCYBR248C/c
LOCUS
DEFINITION
S.cerevisiae chromosome II reading frame ORF YBR248C.
ACCESSION
236117 Y13134
VERSION
236117.1 GI:536662
SOURCE
Saccharomyces cerevisiae (baker's yeast)
ORGANISM
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
REFERENCE
1
AUTHORS
Feldmann,H., Aigle,M., Aljinovic,G., Andre,B., Baclet,M.C.,
Barthe,C., Baur,A., Becam,A.M., Biteau,N., Boles,E., Brandt,T.,
Brendel,M., Brueckner,M., Bussereau,F., Christensen,C.,
Contreras,R., Crouzet,M., Czepluch,C., Demolis,N., Delaveau,T.,
Doignon,P., Dondey,H., Duesterhus,S., Dubois,E., Dujon,B., El
Bakkoury,M., Entian,K.D., Feuerhahn,M., Fiers,W., Fobbo,G.M.,
Fritz,C., Gassenhuber,H., Glanodoff,N., Goffeau,A., Grivell,D.A.,
de Haan,M., Hein,C., Herbert,C.J., Hollenberg,C.P., Holmstrom,K.,
Jacq,C., Jacquet,M., Jauniaux,J.C., Jonniaux,J.L., Kallies,T.,

```

Kiesau, P., Kirchthath, L., Koetter, P., Korol, S., Liebl, S., Logghe, M., Lohan, A.J.E., Louis, E.J., Li, Z.Y., Maat, M.J., Mallet, L., Mannhaup, G., Messenguy, P., Miooga, T., Molemans, P., Mueller, S., Nasr, F., Obermaier, B., Perea, J., Pierard, A., Piravandi, E., Pohl, F.M., Pohl, T.M., Potier, S., Proft, M., Purnelle, B., Ramezani, Rad, M., Rieger, M., Rose, M., Schaaff-Gerstenschlaeger, I., Scherens, B., Schwarzlose, C., Skala, J., Slonimski, P.P., Snits, P.H.M., Souciet, J.L., Steensma, H.Y., Stucka, R., Urestarazu, A., van der Aart, Q.J., van Dyck, L., Vassarotti, A., Vetter, I., Vierendels, F., Vissers, S., Wagner, G., de Weryfossse, P., Wolfe, K.H., Zagulski, M., Zimmermann, F.K., Mewes, H.W. and Klein, K. Complete DNA sequence of yeast chromosome II

EMBO J. 13 (24), 5795-5809 (1994)

95112788
7813418

2 (bases 1 to 2429)
Aljinovic, G., Pohl, F.M. and Pohl, T.M.
Unpublished

3 (bases 1837 to 2429)
Aigle, M., Baclet, M.C., Barthe, C., Biteau, N., Crouzet, M. and Dolignon, F.
Unpublished

4 (bases 1 to 2429)
MIPS.
Direct Submission
Submitted (30-AUG-1994) Data collected by MIPS on behalf of the European yeast chromosome II sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mips.embnat.org
Location/Qualifiers

1. 2429
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/chromosome="II"
complement (352. .2010)
/gene="HIS7"
complement (352. .2010)
/note="ORF YBR248c"
/codon_start=1
/protein_id="CAA85211.1"
/db_xref="GI:536663"
/db_xref="SWISS-PROT:P33734"
/translation="MPVHVIVDSGNLQSLTNAIHLGYEVQLVSPKDFNLSGTSR
LILPGVNGYHFDVNLNFRGPIREYIESGPIGMICVGLQLFAGSVESPKTGL
NYIDFKLSRDFDSEKPYEIAFWNSCIPSENLFFGLDPYKRYFVHSPAILNSEKKQ
LENDGWIKAQYGSSEFIAFVKNKNIATQPHPEKSGKAGLVNENFLKQSPPIPN
VSAPEKLLMNDYSNYGLTRRIIACLDRVNDODLVVTKGDDVDVREKDGKGVNRL
GKPVLAQYKQOGADSVTLNITSFEDCPKDTPELVKQAKTVFVPLTVGGGIK
DIVDVGTKIPALEVASLYFRSGADKVSIGTDVYAAEKYELGNRGDGTSPHETISK
AYGAQAVVISVDPKRVYNSQADTKNVFETPYPGNGEKYCWYQCTIKGGRSRLIG
WVELTRACALGELLNCIDKGSNSGYDLIELIHVDKAVIPVPIASSGAGVPEHP
EEAFKTRADACLAGMFRGEFTVNDVKEVYLLHGLKVRMDEE"

BASE COUNT 714 a 530 c 398 g 787 t

ORIGIN

Query Match 17.2%; Score 336; DB 8; Length 2429;
Best Local Similarity 60.1%; Pred. No. 5.6e-65;
Matches 577; Conservative 0; Mismatches 380; Indels 3; Gaps 1;

Qy 730 TTCTAAACTGCAAGAGAGTAATGCTTGGCTTGATGTGAGGCAACGTAATGATGGGA 789
Db 1312 TTATGGTCTAACACGCGAGATTAATGCTTGTCTTGATGTAGCTACTAATGACCAAGTGA 1253

Qy 790 TCTTGTGTACCAAGGAGACCAATATGATGTGAGAGACGTAACAGAGAGATGAGGT 849
Db 1252 TTTTGGTGTACTAAAGTGATCAATACGATGACGTGAAAAAGTGTATGTAAGGTGT 1193

Qy 850 CAGAAACCTTGGCAAGCTGTGTGAACCTTGTGGGAGGTATTAATTTAGACGGTGTCTGATGA 909
Db 1192 TAGAAACCTTGGTAAGCTGTGTTACGTTGGCAGAGAAATATTACCAACAGGGTGGGATGA 1133

Qy 910 GGTCACTTCTTAAACATTAATCTGGTTTCGGGACATTCCTCTAGGCGATCTACCCATGCT 969
Db 1132 AGTAACATTTTGAATATAACTTCTTTTAGAGATTGCTCTTTGAAGGACTACCGATGCT 1073

Qy 970 ACAGCTTTCGACACGCGCATCTGAAAACGTTTGTGGCCATTAACCTGTGGGGGTGGCAT 1029
Db 1072 AGAGGTTCTGAAACAAGCGCGCAAGACAGCTCTTTGTTCATTTGACAGCTGGTGGGGGAT 1013

Qy 1030 CAGGGATTTTACTGATGCAAAATGGAAGGTATTATTCTAGTCTAGAAAGTGGCTTCAGAGTA 1089
Db 1012 CRAAGATATTGTGATGTTGATGGAACCAAAATACCTGCTTTAGAAAGTGTGCAAGTCTATA 953

Qy 1090 TTTTCAGATCGGCGCGCATAAAGGTTTCGATCGGAAGTGATGCAAGTTTACACTCTCTGAGGA 1149
Db 952 CTTTCAGATCTGGTGTCTGATAAAGTATCGATCGGTACGGATGCAAGTCTATGCAAGGCAAAA 893

Qy 1150 ATATATTAAACCCGAGTGAAGACA---GGAAGAGCAGCATAGAGCAGATATCTACAGT 1206
Db 892 ATACTACGAGTTGGGTAAACAGAGGAGATGGAAGCTCACCATAATAGAGACAAATCTCGAAGC 833

Qy 1207 ATATGCTAACCGAGCGAGTGTGTAAGCATTTGATCTCGCCGAGTTTACTTTGAGAAAAACC 1266
Db 832 ATACCGTGTCTCAGCGAGTTGTTATTCTGTGCGACCTTAAGAGAGTATATGTAATTCACA 773

Qy 1267 CGATGAAGTAGAATTTAAAGCCATCAAGTAAGCCATCCAGGTCCTCAAAACGGTGAGGAATA 1326
Db 772 AGCAGATACGAAGACAAAGTCTTCGAGACAGAAATATCCGGGCCCAATCGAGAGAAATA 713

Qy 1327 TGCCTGGTATCAGTGCACCTGTTAATGGTGAACGAGAGGAGCCCATCGGAGCTTATGA 1386
Db 712 CTGCTGGTACCAATGTACAATCAAGGTGGAAGAGAAATCTAGAGACCTTGTGTGTGGGA 653

Qy 1387 ACTAGCTAAGGCTGTTGAGGAACCTCGAGCTCGAGAAATATTATTTGAACCTGCATTGATTG 1446
Db 652 ATTAACAAGGCGCATGTGAAGCTCTAGGTGCTGGGAGATTTTATTTGAATGTCATAGACAA 593

Qy 1447 TGATGGTCAAGGAAAAAGGATTCGATATAGATCTGATCAAGCTAATATCCGATCTGTGAA 1506
Db 592 GGATGCTCTAATTCCTGTTATGATCTGGAATTTAGTAGAACATGTTAAAGATCGGTCAA 533

Qy 1507 CATTCCTGTTATCGCAAGACGCGTGAGAGTGCCTGATCATTCTTCCGAAAGTCTTTAA 1566
Db 532 GATTCCCGTCAATTCATTCAGTGGCGCGGTGTACCCGAAATTTTCAAGAGGCTTTCCT 473

Qy 1567 TGAACCAACGATCTGCTCCCTTCAGCTGCGCATTTTCCATCGCAAGAGGTTTCCAAT 1626
Db 472 AAAGCCCGCGCAGATGCTTGTGGTGAGGTATGTTCCACAGAGGTGAATTCATGT 413

Qy 1627 TAAGGCTGTTAAAGAGCAGCTTGTGGAAGGAGGATTTGAAGTTAGATTGTAAGGCGGAGAA 1686
Db 412 TAAGATGTAAGGAGTATTTACTAGAGCAGGATTAAGGTTAGAAATGATGAAGAGTA 353

RESULT 6
SCHIS7
LOCUS SCHIS7 2440 bp DNA linear PLN 25-FEB-2003
DEFINITION S.cerevisiae HIS7 gene.
ACCESSION X69815
VERSION X69815.1 GI:395154
KEYWORDS glutamine amido transferase.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
1
REFERENCE
AUTHORS Kuenzler, M., Balmelli, T., Egli, C.M., Paravicini, G. and Braus, G.H.
TITLE Cloning, primary structure, and regulation of the HIS7 gene
encoding a bifunctional glutamine amidotransferase: cyclase from
Saccharomyces cerevisiae
J. Bacteriol. 175 (17), 5548-5558 (1993)
JOURNAL 93374850
MEDLINE PUBMED 8366040
REFERENCE 2 (bases 1 to 2440)

AUTHORS Kuenzler, M.
 TITLE Direct Submission
 JOURNAL Submitted (17-DEC-1992) M. Kuenzler, Mikrobiologisches Institut, Eidgenossische Technische Hochschule, Schmelzbergstrasse 7, ETH Zentrum, 8092 Zuerich, SWITZERLAND

FEATURES
 source Location/Qualifiers
 1. 2440
 /organism="Saccharomyces cerevisiae"
 /mol_type="genomic DNA"
 /strain="S288C congenic"
 /db_xref="taxon:4932"
 /map="Previously mapped, see Mortimer et al. (1992) Yeast 8, 812-902"
 /haplotype="haploid"
 /cell_type="MATA"
 /clone_lib="GENOMIC LIBRARY BASED ON YEAST CENTROMERE PLASMID"
 misc_feature 1. 210
 /note="similarity with E.coli H15H"
 gene 10. 1852
 /gene="HIS7"
 mRNA 10. 1852
 /gene="HIS7"
 /evidence="experimental"
 prim_transcript 10. 105
 /gene="HIS7"
 CDS 106. 1764
 /gene="HIS7"
 /codon_start=1
 /product="glutamine amidotransferase"
 /protein_id="CAA49469.1"
 /db_xref="GI:395155"
 /db_xref="GOA:P33734"
 /db_xref="SCD:S0000452"
 /db_xref="SWISS-PROT:P33734"
 /translation="MPVHVIVDSGNLSLNAIEHLGYEVQLVKSPKDFNIGTSR LILPGVNAHFVDFNIFNRFKPIREYIESGKPMIGICVGLQALFAGVSPKSTGL NYIDFLKSPFDSKEPVPBTGNWSCIPSENLPFLDIPYKRYVPHVFAAILNSEKKK LENDWKLIAKAKYGSBEFIAAVNNKNI FATPHPEKSGKAGLNVI ENFLKQSPPIPN YSAEELKLMNDYSGVLTFRILACLDVNTDQGLVVTKGQDYDVREKSDGKVRNL GRVQLAQRYQOGADEVFLNITSFRDPLKDTPLMLVLEKQNAKTVPVPLTVGGGK DIVDVGTKIPALELYSFRGADKVSIGTDVAYAAEKYELGNRGDGTSPETISK AYGAQVAVISVDPKRYVNSQADTKNKFETEPGPNKGYCWIQCITIKGGRSRLDG VNELTRACALGAGEILLNCIDKGSNSGVDLLEIEHVKDVKIPVIASSGAGVPEHF EEAFLTRADACILGFMFRGEFTVNDVKEYLLEHLGLKVRMDEE"
 mat_peptide 106. 1761
 /gene="HIS7"
 /product="glutamine amidotransferase"
 misc_feature 236. 549
 /gene="HIS7"
 /note="similarity with E.coli H15H"
 BASE COUNT 820 a 405 c 557 g 658 t
 ORIGIN

Query Match 17.2%; Score 336; DB 8; Length 2440;
 Best Local Similarity 60.1%; Pred. No. 5.6e-65;
 Matches 577; Conservative 0; Mismatches 380; Indels 3; Gaps 1;

QY 730 TTCTAAACTTGGCAAGAGTAATTGCTTGCCTTGTATGTGAGGGCAATAATGATTAATGGGA 789
 DB 804 TTATGGTCTAACACGAGAGTAATTGCTTGTCTGTGTACGTACTACTAATGACCAAGGTGA 863

QY 790 TCTTGTGTAAACAGGAGACCAATATGATGTGAGACGCTACAGAGATGAGT 849
 DB 864 TTGTGGTGTACTTAAAGGTGATCAATACGATGTACGTGAAAGAAAGTGGTAAAGGTGT 923

QY 850 CAGAAACCTTGGCAAGCTTCTTCAACTGCTGGCAGTATTATTATAGCGGTCTCATGA 909
 DB 924 TAGAAACCTTGGTAAGCTTGTTCAGTTGGCACAGAAATATTACACAGGTCGGATGA 983

QY 910 GGTGAGCTTTTAAACATTACTGTTTCCGGGACTTCCCTCTAGGCGCATCTACCCATGCT 969
 DB 984 AGTAACATTTTGAATATAACTTCTTTTAGAGATTGTCCTTTGAAGAGTACTCCGATGCT 1043

QY 970 ACAGGCTTTTGAACGCGCATCTGAAACGTTTTTGTGCCATTAACCTGTGGGGTGGCAT 1029
 DB 1044 AGAGGTTCTGAAACAGCGCAAGACAGCTCTTTGTTCATTGACAGCTGGTGGGGGAT 1103

QY 1030 CAGGATTTTACTGATGCAAAATGGAAGGTATTATTCTAGCTTAGAAGTGGCTTCAGAGTA 1089
 DB 1104 CAAGGATATTGTGATGTTGATGGAACCAAAATACCTGCTTTAGAAAGTTGCAAGTCTATA 1163

QY 1090 TTTGAGATCGGCGCGGATAAGGTTTCGATCGAAGTGTGATGAGTTCAGATTTTACACTGCTGAGGA 1149
 DB 1164 CTTGAGATCTGCTGTGATAAAGTATCGATCGTACGATGCAAGTCTTAGACCCGAAA 1223

QY 1150 ATATATTAAACCGGAGTGAAGACA---GGAAAGACGACATAGACAGCATATCTACAGT 1206
 DB 1224 ATACTACGAGTTGGGTAAACAGAGGAGATGGAACGTCAACCAATAGACACAATCTCGAAAGC 1283

QY 1207 ATATGTAACAGGCGAGCTGTTGTAAGCATTTGATCTCTCCGCGAGTTTACTTTGAGAAAACC 1266
 DB 1284 ATACGGTGTCTAGGCGAGTTGTTATTCTCTCGACCCCTAAGAGAGATATATGTAATTCACA 1343

QY 1267 CGATGAAGTAGAATTTAAAGCCATCAAGTAAAGCCATCCAGGTCCTCAAGCGGAGGAGTA 1326
 DB 1344 AGCAGATACGAAGAACAAAGTCTTCGAGACAGAAATATCCGGGCCCAATGGAGAGAAATA 1403

QY 1327 TGCCTGTTATCAAGTGCACCTGTTAAATGGTGGAGAGGAGAGACCCATCGGAGCTTATGA 1386
 DB 1404 CTGCTGTACCAATGTACAAATCAAGGTGGAGAGAAATCTAGAGACCTTGGTGTGGGA 1463

QY 1387 ACTAGCTAAAGCTGTTGAGGAACCTGGAGCTGGAGAAATATTATTGAACTGCATTTGATTG 1446
 DB 1464 ATTAACAAGGCGATGTGAAGCTCTAGGTCTGGGGAGATTTTATTGAACTGCATAGACAA 1523

QY 1447 TGATGTCGAAGAAAGGATTCGATATAGATCTGATCAAGCTAAATCCGATCTCTGAA 1506
 DB 1524 GGATGCTCTAATCTGGTTATGATCTGGAATTTAGTAACATGTTAAAGATCGGTCAA 1583

QY 1507 CATTCCTGTTATCGAAGCAGCGGTGCGAGAGTGCCTGATCACTTCTCCGAACTCTTTAA 1566
 DB 1584 GATTCCTGTCATTCATCAGTGGCGCGGTGTACCCGAACTTTCGAAGAGGCTTCTCT 1643

QY 1567 TGAACCAACAGCATCTGCTGCCCTTGCAGCTGGCATTTTCCATTCGAAAGAGGTTCCAAT 1626
 DB 1644 AAAGACCCGCGAGATGCTTGTCTGGGTGAGTATGTTTCCACAGAGGTGAATTCAGTGT 1703

QY 1627 TAAGCTGTTTAAAGACACTTGTGAGGAAGGATGAGTTAGATTGTAAGCCGAGAA 1686
 DB 1704 TAACGATGTAAAGGAGTATTTACTAGACGCGGATTTAAAGGTTAGAATGGATGAAGATGA 1763

RESULT 7
 NC103EL/c
 LOCUS Neurospora crassa DNA linkage group V Cosmid contig 103EL.
 DEFINITION BX294028
 ACCESSION BX294028.1
 VERSION GI:28950242
 KEYWORDS Neurospora crassa
 SOURCE Neurospora crassa
 ORGANISM Neurospora crassa
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

REFERENCE 1
 Schulte, U., Aign, V., Hoheisel, J., Brandt, P., Fartmann, B., Holland, R., Nyakatura, G., Mewes, H.W. and Mamhouth, G.
 Unpublished
 2 (bases 1 to 87435)
 German Neurospora genome, project.

AUTHORS Submitted (11-MAR-2003) MIPS, Institut fuer Bioinformatik, GSF-Forschungszentrum fuer Umwelt und Gesundheit, GmbH, Ingolstaedter Landstrasse 1, D-85764 Neuherberg, FRG, E-mail: G.Mamhouth@gsf.de
 Project Coordinator: Ulrich Schulte, Institute of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,

JOURNAL
 TITLE
 JOURNAL

		<div><div>/db_xref="GI:28950247"</div><div>/translation="MSDKPSDQTPNQRRSSRRSPGAPVPRDRRDHQRKNTGGA QKAFSLNDQRIIRLLEAQOQCGVGRREPDLERLNQDPSPSPHCFNMARQDSFLK PNAYLAKTITDRLSLALVNOIEKLAMEHTNPVNRVIELVPOFFVEOHRNEVGGTTH DOEPEETALDDOPAPAAVAKAPSVSDVGVGNKCKGRKGTVAQCILIPDPTGLVYG CPICNTDDHLVNCVPKPKFAPGTQLDPKVIQIEYIERLLDVCFVRANRPPFATNGM TWPLFTYRHYAHINTFELNKSMDTGRHYPTWPEYAMRWVDSEVERDEMKKFDP TTTTCNVGCSHSDRPVYNKYESLSETIDAYLDGKWDVATGMIPTTNVDRILTRLMY LEAQRRLVKSVEPEPECAEPAPVPKPSFPPTIVSNTSNLVQVYVWNEETQNF VPDSTDFLSFVTDKAPNLLRCPDEADYATIEYAAVFKHORPONGPVGDSLEDOA LRAVIRAEGLRNRERVAKLGRQAQKRRNAREADREMAFEAQNASTIDVHKREH D"</div></div>			
exon		complement (18190..19685) /gene="103E1.040" /number=1			
intron		complement (19686..19790) /gene="103E1.040" /number=1			
exon		complement (19791..19998) /gene="103E1.040" /number=2			
gene		21036..21715 /gene="103E1.050"			
gene		complement (join(21036..21062,21176..21715)) /gene="103E1.050"			
CDS		complement (join(21036..21062,21176..21715)) /gene="103E1.050" /note="similarity to hypothetical 24.1K protein, Neurospora crassa, PIR:JH0145" /codon_start=1 /product="conserved hypothetical protein" /protein_id="CAD71114.1" /db_xref="GI:28950248"			
<div><div>Query Match</div><div>Best Local Similarity 15.5%; Score 303.2; DB 8; Length 87435;</div><div>Matches 552; Conservative 0; Mismatches 393; Indels 3; Gaps 1;</div></div>					
QY	747	GAGTAATTGCTTGGCTTGATGTGAGGCAAAATGATAATGGGGATCTTGTGTGAACCAAGG	806		
Db	32317	GTGTCATTGCTGCTTGGATGTGGCAACCAACGACGAGGCGACTTGGTGTGTCACCAAGG	32258		
QY	807	GAGACCAATATGATGTGAGAGAACGTACAGAAGAGAAATGAGGTGAGAAACCTTGGCAAGC	866		
Db	32257	GTGACCAATATGATGTCCGCAAAAGTCTGACGACCGCAACGTCGCAACCTCGCAAGC	32198		
QY	867	CTGTTGAACCTTCTGGCGAGTATTATTATAGACGGTGTCTGATGAGGTGAGTCTTCTTAACA	926		
Db	32197	CCGTGAGATGGCGCGCAAGTACTACGAACAAGCGCGCGAAGTGACCTTCTCTCAACA	32138		
QY	927	TTACTGTTTCCGGGACTTCCCTCTAGGCGATCTACCCATGCTACAGGTCTTGCAACGCG	986		
Db	32137	TCACCTCTTCCGCGACTGCCCGTTCGCCACCTGCCATGCTCGAGATCCTTGCCTGA	32078		
QY	987	CATCTGAAAACTTTTGTGCCATTAACCTGTGCGGGGTGGCATCAGGAAATTTTACTGATG	1046		
Db	32077	CCTCAAGACCGCTCTTCGTTCCCTGACCGTCCGCGCGGATCCGCGACACCGTCGACA	32018		
QY	1047	CAAAATGAAGGTATTATTCTAGTCTAGAAAGTGGCTTCAGAGTATTTTCAGATCGGGCGCG	1106		
Db	32017	CGGACGGCACCAAGGTCTCCGCCCTCGAAATCGCCACCATGTACTTCCAGTCAGGCGCGC	31958		
QY	1107	ATAAGTTTCGATCGGAAGTGATGAGTTTTACACTGCTGAGGAATATATATAAACCCGAG	1166		
Db	31957	ACAAGGTGTCCATTGGCTCCGACGCCGTCTATCGTCCCGAGAGTACTATGCTCTCCGCA	31898		
QY	1167	TGAAG---ACAGGAAGAGCAGCATAGACGATATCTTACAGTATATGTTAAACGAGCAG	1223		
Db	31897	AGACGCTCTTTGGCAACACGGCCATTGAGCAGATCAGCAAGGCTACGGAACACGAGCTG	31838		
QY	1224	TGGTTGTAAAGCATTGATCTCTCCGCGAGTTTACTTTGAGAAAAACCGGATGAAGTGAATTTA	1283		
Db	31837	TCGTGCTCTCGGTGATCCCAAGCGGCTCTACGTCCCCAAGCCCGATGCCACGAGACACC	31778		

QY	1284	AAGCCATCAAGTAAGCCCATCCAGGTCCTCAACCGGTGAGGAATATGCTGTATCAGTGCA	1343		
Db	31777	ACACCGTCAAGAGCTGCTACCCCGGCCCAAGSGCGAGGAATACTGCTGTATCGCTGCA	31718		
QY	1344	CTGTTAATGTTGACAGAGAGGAGACCCATCGGAGCTTATGAACCTAGCTAAGCTGTGTTG	1403		
Db	31717	CGATCAAGGCGGCGCGAGACGACGAGGATCTGGATGTGGTTGAACCTCAGCAGGCGGTG	31658		
QY	1404	AGGAACCTTGAGCTGCGAGAAATATTATGAACCTGCATTGATTCTGTATGGTCAAGGAAAAG	1463		
Db	31657	AAGCAATGGCTCGCGGAGATTCTGTTGAATCTGCAATGACAAGGACGACCAACAGCG	31598		
QY	1464	GATTTCGATATGATCTGATCAAGCTAATATCCGATGCTGTGAACATTTCTTGTATCGCAA	1523		
Db	31597	GGTTTGATCTGGAGTTGATCAGGACGTCGAGCGCGCTGAGGATACCGTTATTGCGCA	31538		
QY	1524	GCAGCGGTGACGAGTCTGATCACTTCTCCGAAGTCTTTATGAACCAACGACATCTG	1583		
Db	31537	GCTCGGGGCGGCAACCCGCGGTCACTTTGAGGAGGTGTTTAGGAGACGACACGCGACG	31478		
QY	1584	CTGCCCTTGACGCTGCGCATTTTCCATCGCAAGAGGTTCCTCAATTAAGGCTGTTAAAGAGC	1643		
Db	31477	CCCGCTGGGAGCAGCATGTTCCACCGTGGGAGTATACGTACAGCAGGTCAAGGAGG	31418		
QY	1644	ACTTGTGTAAGGAAGGATTTGAAGTTAGATTGTAAGCGCGAGATCACT	1691		
Db	31417	AGCTGAAGCGGAGGCTGCTGCTCAGACAGTTTTCAGGAGATCTCT	31370		

RESULT 8	AF159463	LOCUS	AF159463	DEFINITION	4351 bp DNA linear PLN 01-JUL-2000
LOCUS	AF159463	DEFINITION	Emericella nidulans glutamine amidotransferase:cyclase (hishF)	gene, complete cds.	
ACCESSION	AF159463	VERSION	AF159463.1	GI:8886089	
KEYWORDS		SOURCE	Emericella nidulans (anamorph: Aspergillus nidulans)	Emericella nidulans	
ORGANISM		REFERENCE	1 (bases 1 to 4351)	Valerius, O., Draht, O., Hoffmann, B., Kuebler, E. and Braus, G. H.	
AUTHORS		TITLE	Isolation and characterization of the hisF gene of Aspergillus nidulans	Unpublished	
JOURNAL		REFERENCE	2 (bases 1 to 4351)	Valerius, O., Draht, O., Hoffmann, B., Kuebler, E. and Braus, G. H.	
AUTHORS		TITLE	Direct Submission	Submitted (16-JUN-1999)	
JOURNAL		FEATURES	Location/Qualifiers	1..4351	
FEATURES		source	/organism="Emericella nidulans"	/mol_type="genomic DNA"	
		gene	/strain="A234"	/db_xref="taxon:162425"	
		mRNA	<519..52329	/gene="hishF"	
		CDS	join(<519..641,728..2176,2240..2329)	/gene="hishF"	
			/product="glutamine amidotransferase:cyclase"	/protein_id="AAF80376.1"	
			/db_xref="GI:8886090"		
			/translation="MPTVHLDDYVAGNVRSLVNAINKVGYEVWVRSPDLKDVEKLI LPGVGHFGHLSQSLSSGYLOPIREHIAKGPKFMGICVGLQSLFSESSEDPNIPGLGT IPARLKRFDKTSVPHIGWNSATDTRIDSTGGQTFYGLSPSSKYYVHVSYAAPYEPG		

ILEKDWLVARRVYBEKEFIGAIARDNIFATQHPBKSGQGRPHSPRFLGRSSAPFC
HPRDSILTKENKGLRRIITACLDVTRNDVGLVTVTKGDVDRKDGADGAGGVORNLG
KPYDMAKKYYEGADEVTEPLNITSFBNCLADLPMLLEILRRTSEVPLVTIGGGIRD
TVDGTHIPALDVASMYFSGADKVSIGSDAVAAEDYYAAGKVLGSKTAIETISKA
YGNQAVVSVDPKRVVSVQPEDTKHRTIETFPNAGQNFQWCTTIKGGRETRDLUDV
COLVQAVEMAGAGEIILNCTDKDGSNSGSIITDQPFESKRAVKIPVITASSGAGMPKHF
EVPDQTTDAALGAGMHRGEYTVGVKQVLEDRGLVRRFEDV"

BASE COUNT 1138 a 975 c 1097 g 1141 t
ORIGIN

Query Match 13.6%; Score 266.6; DB 8; Length 4351;
Best Local Similarity 55.7%; Pred. No. 2.4e-49;
Matches 581; Conservative 0; Mismatches 444; Indels 18; Gaps 3;

QY	578	GAGCATTTTATTCATCATTCAGAGGGAATGTTTCATCGACTCCAAATTCATCCCGAG	637
DB	1133	GAGAAATTCATCGCGCGGATAGCAGAGATAACATCTTTCGACACAAATTCATCCCGAG	1192
QY	638	AAGAGTCGAGGTGTTCGACTTTCATATTCAGAAAGATTTT-----GAATGCTGATTC	691
DB	1193	AAGAGCGGCAGGCGAGCCTAGCACCTTCGCGCTTCTTGACGCGAGCTCAGCTCCA	1252
QY	692	TTTAAACAAAAGACAGAACCAATGAATGGAAGGCTTCTAACTGTCAAAAGAGAGTA	751
DB	1253	TTCTGTACATTTAGAGACTCGATTTTTCAGAGGAGAGAAAACGGTCTTACCCTAGGATC	1312
QY	752	ATTGCTTCCTTGATGTAGGCGCAATGATTAATGCGGATCTTGTGTAAACAGGAGAC	811
DB	1313	ATCGCTGTCTTGATGTTCGTACGAATGATGTGCGGATCTCGTGTGACTAAGGCGCAT	1372
QY	812	CAATATCATGTGAGAGAACGTACAGAAAGAAATG-----AGTTCAGAAACCTTTGGC	862
DB	1373	CAATATGATGTTCGAGAGAGTGTGCGGATGCTGAGGCGCAAGTGAAGAACCTGGGA	1432
QY	863	AAGCCTGTGAACCTGTCTGGGCAAGTATTTTGAACGCTGCTGATGAGGTGACGCTCTTA	922
DB	1433	AAGCGGTGTGATATGCTTAAGAAATATTTACGACAGGGGCGAGATGAGGTGACGTTT	1492
QY	923	AACATTAATGTTTCGCGGACTTCCCTTAGGCGATCTACCCATGCTACAGGTCTTGCAA	982
DB	1493	AACATCACCTCTTTTCAGAAACGTGCTGTTAGCCGACCTCCCTATGCTCGAGATTC	1552
QY	983	CGCCATCTGAAACAGTTTTTGTGCCATTAACCTGCGGGGTGGCATCAGGGAATTTTACT	1042
DB	1553	AGAAGCTCGGAGACCGTCTTCGTAACCTTTCGTAATTTGGTGGCGCATTAGGACACTGTG	1612
QY	1043	GATGCAAAATGGAAGGTATTTCTAGTCTAGAAAGTGGCTTCAGAGTATTTTCAGATCGGG	1102
DB	1613	GATACAGCGGTACTCATACCTCCAGCTCTAGACGTGGCATCGATGTACTTCAAACTGGG	1672
QY	1103	GCCGATTAAGGTTTCGATCGGAAGTATGTCAGTTTACACTGCTGAGGAATATTTAAAC	1162
DB	1673	GCTGACAAAGTCAGCATTTGGTTTCGGATGCGCGTTGCTGCGGAAGATTTATTCGACGT	1732
QY	1163	GG-----AGTGAAGACAGAAAGACAGCATAGACAGATATCTACAGTATATGTTAAAC	1219
DB	1733	GGCAAGTTCTGCTCGGCAAACTGCGCAATGGAACATACTTCTAGGCGGTATGGAAC	1792
QY	1220	GCAGTGGTTGAAGCATTTGATCTTCGCGAGTTTACTTTGAGAAACCCGATGAAGTAGAA	1279
DB	1793	GCTGCTGTTGAAGGTTGACCCGAGCGCGTTTATGTGACCCCAACAGAGACACGAAA	1852
QY	1280	TTTAAAGCCATCAAGTAAGCCATTCAGGTCCAAACCGTGAAGGAATATGCTGATTCAG	1339
DB	1853	CACCGTACGATAGAAACGAAATTTCTTAAACCGCCCGGCAAAATTTCTGTTGTATCCAG	1912
QY	1340	TGCACTGTTATGTTGACGAGAGAGGAGACCCATCGGAGCTTATGAATAGCTTAAGGCT	1399
DB	1913	TGTACTATTAAGGTTGCGAGAGACACGAGACTTAGATGCTGCCAGCTGCTCAAGCC	1972
QY	1400	GTTGAGGAACCTGGAGCTGGAGAAATATTTATGAACTGCAATGATGTTGATGGTCAAGGA	1459
DB	1973	GTCGAGGCAATGGGTGCTGGGAGATTCTGCTGAATTTGCAATGATTAAGACCGGAGCAAC	2032

QY	1460	AAAGGATTCGATATATAGATCTGATCAAGCTAAATATATCGATGCTGTGAACATTCCTGTTATC	1519
DB	2033	AGTGGTTGCGATCTTGACTGATCAACCCAGCGTCAAGCGTGAAGTAAATAATACCCGTGATT	2092
QY	1520	GCAGCAGCGGTGCGAGGAGTCTGATCACTTCTCGAAGTCTTTTAAATGAACCAACGCA	1579
DB	2093	GCCTCTGAGCGAGCTGGTATGCGAAGCAATTTTGAGGAAGTTTTGTGATCAACACGACGACA	2152
QY	1580	TCGTGCTGCCCTTGACGCTGGCAT	1602
DB	2153	GATGCTGCTCTGGGTCTGGGAT	2175

RESULT 9

AC138524.1/c

WPCOMMENT

Sequence split into 6 fragments LOCUS AC138524 Accession AC138524

Fragment Name	Begin	End
AC138524_0	1	110000
AC138524_1	100001	210000
AC138524_2	200001	310000
AC138524_3	300001	410000
AC138524_4	400001	510000
AC138524_5	500001	550277

Continuation (2 of 6) of AC138524 from base 100001 (AC138524 Homo sapiens chromosome 5

Query Match 12.0%; Score 234.8; DB 2; Length 110000;

Best Local Similarity 60.5%; Pred. No. 3.6e-42;
Matches 422; Conservative 0; Mismatches 272; Indels 4; Gaps 2;

QY	730	TTCTAAACTTGCAGAGAGATTAATGCTTTCCTTGTATGTGAGGGCAAAATGATATGGGA	789
DB	21179	TTATGCTCTAACACGAGCAATTAATGCTTGTCTTGTATGCTACTTAATGACCAAGTGA	21120
QY	790	TCTTGTGTAAACGAGGAGACCAATATGATGTGAGAGAACGTACAGAGAGATGAGGT	849
DB	21119	TTTGGTGGTTACTTAAAGGTGATCAATACGATGTACGTGAAAAAAGTATGTTAAAGGTGT	21060
QY	850	CAGAAACCTTGGCAACCTGTTGCAATGCTGGGCGAGTATTATTAGACGGTCTGATGA	909
DB	21059	TAGAAACCTTGGTAAACCTGTTGATGTTGGCAGAGAAATATTACCAACAGGGTGGGATGA	21000
QY	910	GGTCAGCTTCTTAAACATTAATGCTGTTTCCGGGACTTCCCTCTAGGGCATCTACCCATGCT	969
DB	20999	AGTAACATTTTGAATATAACTTCTTTTAGAGATTGCTCTTGAAGGATATCTCCGATGCT	20940
QY	970	ACAGGTCTTGCAACGCGCATCTGAAAAAGTCTTGTGCGCATTAACCTGTCGGGGTGGCAT	1029
DB	20939	AGAGGTTCTGAAACAAAGCCGCAAGACAGTCTTTGTTCATTTGACAGTCCGTGGGGGGAT	20880
QY	1030	CAGGGATTTTACTGATGCAAAATCGAAGGTATTATTCTAGTCTAGAGTGGCTTCAGAGTA	1089
DB	20879	CAAGGATATTGTCGATGTTGATGGAACCAAAATACCTGCTTTAGAGTCCGAGTCTATA	20820
QY	1090	TTTCAGATCGGGCGCGGATAAGGTTTCGATCGGAAGTATGTCAGTATTACACTGCTGAGGA	1149
DB	20819	CTTCAGATCTGGTCTGATTAAGTATCGATCGGTACGATGCGATGTCAGTCTATCAGCCGAAA	20760
QY	1150	ATATATTAAACCGGAGTGAAGACA---GGAAGACGACGATAGAGCAGATATCTACAGT	1206
DB	20759	ATACTATGAGTGGGTAAACAGAGAGATGGAACGCTCAACCAATAGAGCAATCTCGAAAGC	20700
QY	1207	ATATGTTAAACGAGGAGTGGTTCGTAAGCATTTGATCTCGCGGAGTTTACTTTGAGAAAAAC	1266
DB	20699	ATACGGTGTCTCAGCAGTGTGTTATTTCTGCGACCTTAAGAGAGTATATGTCATTCACA	20640
QY	1267	CGATGAAGTGAATTTTAAAGCCATCAAGTAAGCCATCCAGGTCCAAACCGGTGAGGAATA	1326
DB	20639	AGCAGATACCAAGAACAAACGCTCTTCGAGACAGAAATATCCGGGCCCAATCGAGAGAAATA	20580
QY	1327	TGCTGTGT-ATCAGTGCACGTGTTAATGCTGACGAGAGAGGAGACCCATCGAGCTTATG	1385

Db 20579 CTGCTGGTCTCAATGTGCAATCAATGGTGAAGAGAAATCTAGAGACCTTGGTGTGGG 20520

QY 1386 AACTAGTAAGGCTGTGTGAGAACTTGGAGCTGGAGAA 1423

Db 20519 AATTAACAGGGCATGCGAAGCTCTAGTCTGGGGA 20482

RESULT 10
CNS06K2Q/c 975 bp DNA linear STS 10-JAN-2001
LOCUS T3 end of clone AT0AA001802 of library AT0AA from strain CBS 4311
DEFINITION of Saccharomyces servazzii, sequence tagged site.

ACCESSION AL402360
VERSION AL402360.1 GI:12161223

KEYWORDS Saccharomyces servazzii
SOURCE Saccharomyces servazzii
ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

REFERENCE 1 (bases 1 to 975)
AUTHORS Souciet J.L., Aigle M., Artiguenave F., Blandin G., Bolotin-Fukuhara M., Bon E., Brottier P., Casaregola S., de-Montigny J., Dujon B., Durrens P., Lepingle A., Llorente B., Maupertuy A., Neuveglise C., Ozier-Kalogeropoulos O., Potier S., Saurin W., Tekala F., Toffano-Nioche C., Wesolowski-Louvel M., Wincker P. and Weissenbach J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876

REFERENCE 2 (bases 1 to 975)
AUTHORS Casaregola S., Lepingle A., Bon E., Neuveglise C., Nguyen H., Artiguenave F., Wincker P. and Gaillardin C.
TITLE Genomic exploration of the hemiascomycetous yeasts: 7.

JOURNAL Saccharomyces servazzii
MEDLINE FEBS Lett. 487 (1), 47-51 (2000)
PUBMED 20584717

REFERENCE 3 (bases 1 to 975)
AUTHORS Neuveglise C., Lepingle A., Bon E., Artiguenave F., Wincker P., Dujon B., Durrens P., de-Montigny J., Maupertuy A., Saurin W., Tekala F., Toffano-Nioche C., Wesolowski-Louvel M., Wincker P. and Weissenbach J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL Saccharomyces servazzii
MEDLINE FEBS Lett. 487 (1), 47-51 (2000)
PUBMED 20584717

REFERENCE 4 (bases 1 to 975)
AUTHORS Neuveglise C., Lepingle A., Bon E., Artiguenave F., Wincker P., Dujon B., Durrens P., de-Montigny J., Maupertuy A., Saurin W., Tekala F., Toffano-Nioche C., Wesolowski-Louvel M., Wincker P. and Weissenbach J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL Saccharomyces servazzii
MEDLINE FEBS Lett. 487 (1), 47-51 (2000)
PUBMED 20584717

REFERENCE 5 (bases 1 to 975)
AUTHORS Neuveglise C., Lepingle A., Bon E., Artiguenave F., Wincker P., Dujon B., Durrens P., de-Montigny J., Maupertuy A., Saurin W., Tekala F., Toffano-Nioche C., Wesolowski-Louvel M., Wincker P. and Weissenbach J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL Saccharomyces servazzii
MEDLINE FEBS Lett. 487 (1), 47-51 (2000)
PUBMED 20584717

REFERENCE 6 (bases 1 to 975)
AUTHORS Neuveglise C., Lepingle A., Bon E., Artiguenave F., Wincker P., Dujon B., Durrens P., de-Montigny J., Maupertuy A., Saurin W., Tekala F., Toffano-Nioche C., Wesolowski-Louvel M., Wincker P. and Weissenbach J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL Saccharomyces servazzii
MEDLINE FEBS Lett. 487 (1), 47-51 (2000)
PUBMED 20584717

REFERENCE 7 (bases 1 to 975)
AUTHORS Neuveglise C., Lepingle A., Bon E., Artiguenave F., Wincker P., Dujon B., Durrens P., de-Montigny J., Maupertuy A., Saurin W., Tekala F., Toffano-Nioche C., Wesolowski-Louvel M., Wincker P. and Weissenbach J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL Saccharomyces servazzii
MEDLINE FEBS Lett. 487 (1), 47-51 (2000)
PUBMED 20584717

REFERENCE 8 (bases 1 to 975)
AUTHORS Neuveglise C., Lepingle A., Bon E., Artiguenave F., Wincker P., Dujon B., Durrens P., de-Montigny J., Maupertuy A., Saurin W., Tekala F., Toffano-Nioche C., Wesolowski-Louvel M., Wincker P. and Weissenbach J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL Saccharomyces servazzii
MEDLINE FEBS Lett. 487 (1), 47-51 (2000)
PUBMED 20584717

REFERENCE 9 (bases 1 to 975)
AUTHORS Neuveglise C., Lepingle A., Bon E., Artiguenave F., Wincker P., Dujon B., Durrens P., de-Montigny J., Maupertuy A., Saurin W., Tekala F., Toffano-Nioche C., Wesolowski-Louvel M., Wincker P. and Weissenbach J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL Saccharomyces servazzii
MEDLINE FEBS Lett. 487 (1), 47-51 (2000)
PUBMED 20584717

Matches 437; Conservative 0; Mismatches 333; Indels 3; Gaps 1;

QY 905 GATGAGGTGAGTCTTCTTAAACATTACTGCTTTTCGGGACTTCCCTCTAGCCGATCTACCC 964

Db 974 GATGAAATTACATCTTCTTAAATATACCTCTTTTAGAGATTSCCATTAAGATATCTCCA 915

QY 965 ATGCTACAGGTCTTGCAACGGCATCTGAAACGTTTTTGTGTCATTAACTGTCGGGGT 1024

Db 914 ATGTTGGAATATTAAGATTGGCGCTAAGACAGTTTTTCTGTTCCCTTGACTGTGGCGGT 855

QY 1025 GGCATCAGGATTTTACTGATGCAATGCAAGGTATTTACTAGTCTAGAGTGGCTTCA 1084

Db 854 GGCATTAAGATATCATTTGATATGACGGAACCAAGATTCCTGCTTTAGACGTTGGTAGT 795

QY 1085 GAGTATTTTCAGATCGGGCGCCGATGAGTTTCGATCGAAGTATGATGAGTTTACACTGCT 1144

Db 794 TTATATTTTATGATCTGCTGCGGATTAAGTTTCAATTTGGTACAGTCCGCTGATGCTGT 735

QY 1145 GAGGAATATATTAACACCGAGTGAAGACA---GGAAAGAGCAGCATAGACGAGATATCT 1201

Db 734 GAAATGTTTATCTTAACGCTGCAAGAGGAGACGGTCTTCCCAATTGAAACCATCTCA 675

QY 1202 ACAGTATATGTAACACCGAGTGGTGTGAAGCATTTGATGCTCCGCGAGTTTACTTTGAGA 1261

Db 674 AAGGCATATGGAGCACAAGCTGTGGTTATTTCTGTTGATCCAAAGAAAGTTTATGTAAT 615

QY 1262 AAACCCGATGAAGTAGAATTTAAAGCCATCAAGTAAGCATCCAGGTCCTCAACCGGTGAG 1321

Db 614 GATCCAAGAGATACCAAGAAATTAAGCATTTTAAAGTAAAGTATGCTCCGATTTCAAAGGGTGA 555

QY 1322 GAATATGCTGCTGATCAGTGCATCTGTTAATGTTGACGAGAGAGGAGACCCATCGGAGCT 1381

Db 554 ACATGCTGTTGTTATCAATGTCACCATCAAGGTTGGTAGGAGTCAAGAGATTTAGGTGT 495

QY 1382 TATGAAGTATGAGTAAAGCTGTTGAGGAACTTGGAGCTGGAGAAATATTTTGAATGCTAT 1441

Db 494 TGGGAATTAATTCAGCTTTGTAAGCCCTTGGGCGAGGTGAATTCCTTTGAAATTTGATT 435

QY 1442 GATTGTGATGCTCAAGGAAAGGATTCGATATAGATCTCATCAAGCTAATATCCGATGCT 1501

Db 434 GATAAGACGGGTCTTAATTCAGTTATGATCTGGAATTAATCGAGCAAGCAAGGATGCA 375

QY 1502 GTGAACATTTCTGTTATCGCAAGCAGCGGTGCGAGGAGTCCGTCATCACTTCTCCGAAGTC 1561

Db 374 GCGAAGATCCAGTCATTCCTCGAGTGTGCTGTTAAACAGCTCATTTTCGAAGAAGCT 315

QY 1562 TTTAATGAACCAACGATCTGCTGCCCTTGGAGCTGGGATTTTCCATCGCAAGAGGTT 1621

Db 314 TTTAGAAAAACCAAGAGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 255

QY 1622 CCAATTAAGCTGTTAAAGAGCAGCTTTGTAAGGAGGAGGATTCGAGTTAGATT 1674

Db 254 ACGTTAGTACGTGAAGATTTATTTACTACACCAATTTTCAGAGTAAGAAT 202

RESULT 11
AF290177
LOCUS AF290177
DEFINITION Candida albicans imidazole glycerol phosphate synthase (HIS7) gene, complete cds.

ACCESSION AF290177
VERSION AF290177.1 GI:10444338

KEYWORDS Candida albicans
SOURCE Candida albicans

ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE 1 (bases 1 to 2306)
AUTHORS Day T.W. and Davison V.J.

TITLE Cloning and Characterization of the CaHIS6 and CaHIS7 Genes from the Fungal Pathogen Candida albicans

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2306)

Query Match 11.5%; Score 224.2; DB 11; Length 975;
Best Local Similarity 56.5%; Pred. No. 8.4e-40;

BASE COUNT 323 a 195 c 149 g 307 t 1 others

misc_feature complement (<198. .5974)
/notes="similar to Saccharomyces cerevisiae ORF YBR248c [HIS7 ; glutamine amidotransferase/cyclase]"
/evidence=not experimental

AUTHORS Day, T. W. and Davison, V. J.
 TITLE Direct Submission
 JOURNAL Submitted (26-JUL-2000) Medicinal Chemistry and Molecular Pharmacology, Purdue University, 1333 Robert Heine Pharmacy Building, West Lafayette, IN 47907-1333, USA

FEATURES
 source 1. .2306
 /organism="Candida albicans"
 /mol_type="genomic DNA"
 /db_xref="taxon:5476"
 <161. .>2044
 /gene="HIS7"
 <161. .>2044
 /gene="HIS7"
 /product="imidazole glycerol phosphate synthase"
 .161. .2044
 /gene="HIS7"
 /note="class I glutamine amidotransferase; IGP synthase"
 /codon_start=1
 /transl_table=12
 /product="imidazole glycerol phosphate synthase"
 /protein_id="AAGI7872.1"
 /db_xref="GI:10444339"
 /translation="MTKTHIIVESGNLSNAIKRIDSNIYIKFPIHNEQDFDND
 SOIKLIFPGVGNDFHFKOLNERKLINHLKSVIKODRPLMGI CVGLQSI FHESESP
 NIKGLGLDNDDDDEKLLKFKDNDDEKFKIRGIKKSVPHIGWNNIHDIIING
 KQTTSLYGLNKIDKYFVHSYAAIIKNNDEIIENKFKIKLSSKGMOLAIQVGE
 KFIATSKNNLFPKPEKSGIVGLKIIKFNLEKNPPTPTPTFNNAQGVDS
 SSTLDVETTLGTIRYIIACLDVNTDGLVTVKQDQNVNREQLSSSTNESABEDS
 SOVNLKPEVLATKYVNOGADBITFLNITSFNSPLKDLPLMLQVLKAAETIFVPLT
 VGGIKDLYDPETKVPVAVDAHLVFSQADKISIGSDAVTIAENVYANNIGITGLS
 SIETISKFGNOAVTVSDPKRYIKDPTTETTTMOTIKITDPSQVNGEYOVCYQV
 TSQGRKIHGALBELCLACEKLGABEILLNSIDHDSNKGPNLQLOLQIKSVISIPV
 IASSGAGNPQHFQVFMDCGIDALGAGLHFRGEYTVNQVKYLOQEAQMDVRLED
 LEL"

BASE COUNT 844 a 327 c 376 g 759 t
 ORIGIN

Query Match 10.1%; Score 198.2; DB 8; Length 2306;
 Best Local Similarity 54.9%; Pred. No. 6e-34;
 Matches 535; Conservative 0; Mismatches 383; Indels 57; Gaps 5;

Qy 741 CAAGAGAGTAATGCTTGCCTCATGTGAGGCAAAATGATAATGAGGATCTTGTGTAA 800
 Db 1029 CTAGAAGATTTGTCATGCTTGTGATGTAGAACTAATGATGATGATGATGATG 1088
 Qy 801 CCAAGGAGGACCAATATGATGTGAGAAACGTACA----- 835
 Db 1089 CTAAGGTGATCAATAATATGTTCCGAGAACAAATTCATCATCATCACTAAACGAATCTG 1148
 Qy 836 -----GAAGAAATGAGTCAAGAACCTTGGCAAGCCTGTGAACTTGTCTGGGCAGT 887
 Db 1149 CCGAGGAGGATTTCTCAGGTTCTGTAATTTAGGTAACACGATTTGAATAGCAACAAAT 1208
 Qy 888 ATATTTTAGACGGTGTGATGAGTCACTTCTTAAACATTTACCTGTTCCGGGACTTC 947
 Db 1209 ATTATATCAAGGAGCTGATGAATATACATTTTAAATATACATCTTTCGTAATCTC 1268
 Qy 948 CTCTAGGCGATCTACCATGCTACAGGTCTTGCAACCGCATCTGAAACGTTTTTCTGTC 1007
 Db 1269 CTTTAAAGATTTACCATGTTTCAAGTATTGAAAGAAAGCTGCTGAACTATATTCTTC 1328
 Qy 1008 CATTAAGTCTGGGGTGGCATCAGGATTTTACTGAT---GCAATAGGAGGTATTATT 1064
 Db 1329 CATTCAGTGTGGTGGAGGTATTAAAGATTTATATGATCCAGAACTGGGAAACCTGTC 1388
 Qy 1065 CTAGTCTAGAAGTGGCTTCAGATGATTTTCAGATCGGCGCGGATGAAAGTTTCGATCGGAA 1124
 Db 1389 CTGCTGTGATGTGGCTCATTTATTTTCCAACTCAGGTGCTGATAAATAGTATGAT 1448
 Qy 1125 GTGATCAGATTTTACACTGCTGAGGAATATATT---AAACCGGAGTGAAGACAGGAAGA 1181
 Db 1449 CTGATCGCGTTACCATTGCCGAAATATTATTATGCAATATATGCGGAAATATACCGGCTCTT 1508

Qy 1182 GCAGCATAGACAGATATCTACAGTATATGTTAAACAGGACAGTGTGTTAAGCATTTGATC 1241
 Db 1509 CATCAATTTGAACCACTTCTTCCAAATTTGGTAAATCAAGCAGTGTGTTATATCAGTTGATC 1568
 Qy 1242 CTCGCCGAGTTTACTTTGAGAAAC-----CCGATGAAGTAGAATTTAAAG 1286
 Db 1569 CTAAGAAATAATATCAAGATCCAACTGAACTACTACTATGCAACAAATAAAAA 1628
 Qy 1287 CCATCAAGTAAGCCATCCAGGTCCAAACGGTGAGAAATATGCTTGGTATCAGTGCACATG 1346
 Db 1629 TCATCGATCCTTTCACAAATATGGTCCAAATGGTGAACAATATTGTTATTATCAAGTAACCT 1688
 Qy 1347 TTAATCGTGCAGAGAGGAGACCCATCGAGCTTATGAACCTAGCTTAAGGCTGTTGAGG 1406
 Db 1689 CTCAGGTGAGAGAAATTCATGAATAGGTGCATTAATATTTAGCTTGTGAAA 1748
 Qy 1407 AACTTGGAGCTGGAGAAATATTATTGAACCTGATGATGATGATGATGATGATGATGATGATG 1466
 Db 1749 AATTGGGAGCTGGAGAAATTTTATTAAATTCATCGATCATGATGATGATGATGATGATGATG 1808
 Qy 1467 TCGATATAGATCTGATCAAGCTAAATATCCGATCTGTGAACATTCCTGTTATCGCAGCA 1526
 Db 1809 TTAATTTTACAAATTTATTACAACTAAATCTCAAGTTTCAATCCCGGATTCGATGATG 1868
 Qy 1527 GCGTGCAGAGTCGCTGATCACTTCTCGAAAGTCTTT---AATGAAACCAACGCATCTG 1583
 Db 1869 GTGGTCTGTAATCTCAACATTTCCACAGATTTTGAATGATTTGGAATTTGATG 1928
 Qy 1584 CTGCCCCTTCAGCTGCATTTTCCATCGCAAGAGGTTTCCAAATGAAGCTGTTTAAAGAGC 1643
 Db 1929 CTGCTTTAGTGTGCTGTTTTCATCGTGGTGAATATACCGTTAATCAAGTAAGAAAT 1988
 Qy 1644 ACTTGTGAAGGAG 1658
 Db 1989 ATTTACAACAGGAAG 2003

RESULT 12

AB016783
 LOCUS Arabidopsis thaliana 4581 bp DNA linear PLN 06-FEB-1999
 DEFINITION partial cds.
 ACCESSION AB016783
 VERSION 1 GI:3413533
 KEYWORDS glutamine amidotransferase/cyclase.
 SOURCE Arabidopsis thaliana (thale cress).
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (sites)
 Fujimori, K. and Ohta, D.
 An Arabidopsis cDNA encoding a bifunctional glutamine
 amidotransferase/cyclase suppresses the histidine auxotrophy of a
 Saccharomyces cerevisiae his7 mutant
 FEBS Lett. 428 (3), 229-234 (1998)
 JOURNAL MEDLINE 98316711
 PUBMED 9654139
 REFERENCE 2 (bases 1 to 4581)
 Fujimori, K.
 Direct Submission
 Submitted (05-AUG-1998) Ko Fujimori, Osaka Bioscience Institute,
 Department of Molecular Behavioral Biology; 6-2-4 Furuedai, Suita,
 Osaka 565-0874, Japan (E-mail:fujimori@obi.or.jp,
 Tel:+81-6-872-4851, Fax:+81-6-872-2841)
 Location/Qualifiers
 1. .4581
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /db_xref="taxon:3702"
 join(2069. .2255,2348. .2430,2524. .2618,2711. .2814,
 2928. .3212,3313. .3407,3521. .3619,3800. .3862,3970. .4032,
 CDS

```
4125. .4207,4338. .4446,4547. .>4581)
/codon_start=1
/product="glutamine amidotransferase/cyclase"
/protein_id="BAJ32287.1"
/db_xref="GI:3413534"
/translation="MENAAPFSSIVSRNFFSSSSIRASSPASLFLSKSIGNNVR
KFSPRLSVRASTSDSVTLDDYAGNVRISRNALRHIGFSIKDQVTPGDIILNADR
LIFPGVPFAPADVLNRMTGMAELCKYIENDRPFILGICLQLLDFSSQNGPVRGL
GVIPGIVGRFDAGIRVPHIGWALQVGDSDILDDVGNRHVVFVHSVRAIPSDENK
DWSSTCNVGEISFISRRGNVHAVOPHPKSGEGLSVLRFLPKLPATOKPMGSK
ASKLAKVIACLDVNTDKGLVVTGDDVREQSNENRNLGKPVDLAGVYKDG
ADSEFINITGRDFPLGDLPMQLQVLRQTSKNVFPVLTGGRDFDASGRYISLE
VAEIFKSGADKPSIGSDVFAAEEFKSGVTKGKSLEIS"
BASE COUNT      1349 a   782 c   816 g   1634 t
ORIGIN
      778;   Score 150.8;   DB 8;   Length 4581;
      Best Local Similarity 66.8%;   Pred. No. 2.8e-23;
      Matches 231;   Conservative 0;   Mismatches 112;   Indels 3;   Gaps 1;

Qy      324  GATTGACGACTCTTTTGAGTCAAGTGAAGAAATGCTCAATTCAGAGTCTTGGCTTGA 383
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      2886  GATATCATGTATTTTGTACTTAACTATGATGTGACAGTCAAGGTCTTGGTGTGA 2945

Qy      384  TTCTGACGCGGTTGGCGTTTGAATCATCCAAATGTTTAAAGGTGCCACATATTGGAT 443
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      2946  TACCGGAATAGTTGGACGCTTTGATGCTTCAGCTGTATATAAGAGTACCCACATGGCT 3005

Qy      444  GGCATGCTTGGATATAAGGAAGGTTCAGCAATTTTAGATGATGTGGGAATCAACATG 503
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      3006  GGAATGCTTTGCAAGTTGGGAAGGATTTCTGAAATTTTGGATGATGTTGGAAACCGTCATG 3065

Qy      504  TGTATTTTGTCTACTCATATCGACCAAT---GCCGAGGACACAAAGAGTGATGATTTTCAT 560
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      3066  TCTATTTTGTTCATCTGACAGGCCATTCATCAGATGAATAAGACTGATGATTTGCT 3125

Qy      561  CTACATGACGCTATGTCAGCAATTTTATTCATCCATTTCAGAGGGAATGTTTCATGACG 620
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      3126  CTACCTGTAATTTGGTGAATCATTTTATATCTTCCATAAGAGGGGAATGTGATGACG 3185

Qy      621  TCCAAATTCATCCGAGAGAGTGAGGAGTGTGGACTTTCATTT 666
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      3186  TTCAATTCATCTCTGAAAGAGCGGGGGTATGATAGTCTAGTCATT 3231

RESULT 13
AC134234/c
LOCUS      AC134234      195583 bp   DNA   linear   PLN 11-MAR-2003
DEFINITION      Oryza sativa (japonica cultivar-group) chromosome 3 clone
ACCESSION      OSJNBa0019J12, complete sequence.
VERSION      AC134234.3   GI:28913003
KEYWORDS      HTG.
SOURCE      Oryza sativa (japonica cultivar-group)
ORGANISM      Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE      1. (bases 1 to 195583)
AUTHORS      Wing,R.A., Yu,Y., Soderlund,C., Kim,H.-R., Rambo,T., Currie,J. and
            Collura,K.
TITLE      Rice Genomic Sequence
JOURNAL      Unpublished
REFERENCE      2. (bases 1 to 195583)
AUTHORS      Wing,R.A., Yu,Y., Soderlund,C., Kim,H.-R., Rambo,T., Saski,C.,
            Currie,J. and Collura,K.
TITLE      Direct Submission
JOURNAL      Submitted (24-SEP-2002) Arizona Genomics Institute, University of
            Arizona, 303 Forbes, Tucson, AZ 85721, USA
REFERENCE      3. (bases 1 to 195583)
AUTHORS      Wing,R.A., Yu,Y., Soderlund,C., Kim,H.-R., Rambo,T., Saski,C.,
            Currie,J. and Collura,K.
TITLE      Direct Submission

Submitted (23-OCT-2002) Arizona Genomics Institute, University of
Arizona, 303 Forbes, Tucson, AZ 85721, USA
REFERENCE      4. (bases 1 to 195583)
AUTHORS      Wing,R.A., Yu,Y., Soderlund,C., Kim,H.-R., Rambo,T., Currie,J. and
            Collura,K.
TITLE      Direct Submission
JOURNAL      Submitted (11-MAR-2003) Arizona Genomics Institute, University of
            Arizona, 303 Forbes, Tucson, AZ 85721, USA
COMMENT      On Mar 11, 2003 this sequence version replaced gi:24270658.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality
            >30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by more than
            one plasmid subclone; and the assembly was confirmed by
            restriction digest. The nucleotide sequence of this BAC clone was
            generated by combining Syngenta, Monsanto and Arizona Genomics
            Institute sequencing data. Between the bases at 64851 and 64854
            there may be two Ts in a run of Gs and between the bases at 66163
            and 66164 there may be an A in a run of Cs. There is only Syngenta
            reads at 88297-88394 and only Monsanto reads at 111365-111557.
            At bases 40328-40344 there is only coverage from PCR reads.
            Bacterial Transposons can be found at the following
            locations: 61245-63667. The assembly overlaps from base 1-7365 with
            OSJNBb0081110 (accession # AC134240). The overlap is from
            135933-143298 bases on OSJNBb0081110.
            Location/Qualifiers
                1..195583
                /organism="Oryza sativa (japonica cultivar-group)"
                /mol_type="genomic DNA"
                /db_xref="taxon:39947"
                /chromosome="3"
                /clone="OSJNBa0019J12"
                /base_count 53686 a 44562 c 44555 g 52780 t
BASE COUNT      53686 a 44562 c 44555 g 52780 t
ORIGIN
      6.6%;   Score 129.6;   DB 8;   Length 195583;
      Best Local Similarity 64.9%;   Pred. No. 1.7e-18;
      Matches 192;   Conservative 0;   Mismatches 104;   Indels 0;   Gaps 0;

Qy      81  CTGTGTGACTTTACTTTGATTACGGTCTGGAAATGTTAGGAGTGTGAGAACGCCATCC 140
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      162857  CAGCCGTGACGTGCTGGACTACGGCGCGGCAACGTGCGCAGCGTGCAGCATCCATCC 162798

Qy      141  GCACACTTCGATTTGATATCAAGATGTCAAAGCCAGAGGATATCTTAAATGCTAAGC 200
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      162797  GCACCTCGGCTTCGGCATCCGGACGTGGCAGCCGCGGAGACATCTTCGCCGCCGACC 162738

Qy      201  GCCTTATCTTCTCGCGTGGGGCTTTGACACCTGCAATGGATGTTCTTATTTCGTAAG 260
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      162737  GCCTCGTCTTCCGGGGGTGCGGCGCTTCGGCTCAGCCATGACGCTCTCACCGCTCCG 162678

Qy      261  GGCTGCTGAAGCATTCTGTACTTACATTCAGAAATGATCGACCTTCTCGGGTATATGCC 320
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      162677  GGATGCGCCAGCGCTCCCGAGTATCATCCGAGGACGCCGCCCTTCTCTCGCATCTGCC 162618

Qy      321  TCGGATTGACGACTCTTTGACTCAAGTGAAGAAATGTCCTCAATTCAGGTCCT 376
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      162617  TCGGCTCCAGCTTCTCTTCGACTCTAGCAGAGAAATGCGCCGAGTAAAGCTCCT 162562

RESULT 14
AL591975/c
LOCUS      AL591975      280050 bp   DNA   linear   BCT 06-JUN-2002
DEFINITION      Listeria monocytogenes strain EGD, complete genome, segment 3/12.
ACCESSION      AL591975 AL591824
VERSION      AL591975.1   GI:16409709
KEYWORDS      Listeria monocytogenes
SOURCE      Listeria monocytogenes
ORGANISM      Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE      1
AUTHORS      Glaser,P., Frangeul,L., Buchrieser,C., Rusniok,C., Amend,A.,
```

Baquero, F., Berche, P., Bloecker, H., Brandt, P., Chakraborty, T., Charbit, A., Cherouani, F., Couve, E., de Darvar, A., Dehoux, P., Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L., Dussurget, O., Entian, K.D., Feihl, H., Portillo, F.G., Garrido, P., Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J., Jackson, D., Jones, L.M., Kaerst, U., Kretz, J., Kuhn, M., Kunst, F., Kurapkot, G., Madueno, E., Maitounam, A., Vicente, J.M., Ng, E., Nedjari, H., Nordstiek, G., Novella, S., de Pablo, B., Perez-Diaz, J.C., Purcell, R., Rammel, B., Rose, W., Schlueter, T., Simoes, N., Tierrez, A., Vazquez-Boland, J.A., Voss, H., Wehland, J. and Cossart, P. Comparative genomics of *Listeria species* Science 294 (5543), 849-852 (2001)

21537279
11679669

2. (bases 1 to 280050)
Glaser, P., Frangeul, L. and Rusniok, C.
Direct Submission
Submitted (06-JUN-2001) Glaser P., Institut Pasteur, Genomique des Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE

E-mail: pglaser@pasteur.fr
Phone: +33 1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.

FEATURES

source

1. 280050
/organism="Listeria monocytogenes"
/mol_type="genomic DNA"
/strain="EGD-e"
/db_xref="taxon:1639"

gene

158..584
/gene="lmo0332"

RBS

158..163
/gene="lmo0332"

CDS

172..507
/gene="lmo0332"

/codon_start=1
/transl_table=11
/protein_id="CAC98411.1"
/db_xref="GI:16409710"

/db_xref="SPTREMBL:O8YA33"

/translation="MIYICALYTFISALVSFGFSLDALLSKRVNGDALINKAYAVS
RSLSLVALGLFLFKSDAFVALSLVMIGALQFDGIGIKISTFKTVGPLLTAAGNV
IMLLFLTI"

terminator

565..584
/gene="lmo0332"

936..6272
/gene="lmo0333"

936..6272
/gene="lmo0333"

/note="similar to internalin proteins, putative
peptidoglycan bound protein (LPXTG motif)"
/codon_start=1
/transl_table=11
/protein_id="CAC98412.1"
/db_xref="GI:16409711"

/db_xref="SPTREMBL:O8YA32"

/translation="MKKFSIVISVLLGLVLPFDLLVCADETTVSEDVAVKTAEA
DSATEGLESTGDETAEPKEAEAEAKETTEREKAKTEEPASNIKTEINDKS
OLQKTSKAAVPAGSTYNSLPDNDLAKLAVITGNAAATGNESVDSAAALAIQLD
LSGTGNDPTDISNIEGLQYLDNLSLNNISDLAPKLQVNLVSLNLSNRILV
NLGVEDLVNQLVNSANKALEDSQVSLPVLKEISAOQCNKITLEKNPAGAVLP
ELETFLQNDLNLTSKLPLKLVYKGNLSKLETLNGATKLQIDASNCITD
ETLDIGLSELEMQLSGSKLKEITSLKNLNLVNIADSCAIEDLGTNLLENK
TLVSDNENLTNITDLPQKLTLDGGCITSIGTLDNLPLEKLDLKENQITSI
EITDPLSLVDSVNNLTITDGLKPLLELWVNSNRSLSDVSLTNFPSLNYNIS
NNVIRTVKSNRLESTKFEYAQNNSISDI SMIHDPNLRKVDASNNLTINIGFDNLP
KLQSLDVHNSRTSLSTVHDLPSLETFAOQTNLTIGTMDNLPDLYVNLNFRNIPS
LAPIGDPLNLTIVSDNNSVLSRISGTMDGVKRLIIDLQNNYLVNTGTGECNLSLSD
LTNLTSLNLRNVVIDISGLTSLRLIYLNLSNKIEDISALSNLTNLOELTLNNK
IENISLSDLENLNLVSVKNNIIDSFVANNVRGAIVTASQTYTLPTVLSYQSSP
TIDNPVIAWYDGLTLPASSIGNSGNYKDKITWTNNATSSSLTFNFRNLKDLTFSGT
VTPYKAAKVTADEQTYTIGDTISEEQFKOVNAKSSDGPVTFATVNDITFG
EYEVTLTSEKDIGQSKVIVKLVHGAPVISADQTIISYDKHATITTEKQLEDIHAST
DLDTAITTNFSTAVLNKGGDYTVALNSENEDGVKAETVYTVTVNKKDPAIISAKTE

/transl_table=11
/protein_id="CAC98411.1"
/db_xref="GI:16409710"

/db_xref="SPTREMBL:O8YA32"

/translation="MKKFSIVISVLLGLVLPFDLLVCADETTVSEDVAVKTAEA
DSATEGLESTGDETAEPKEAEAEAKETTEREKAKTEEPASNIKTEINDKS
OLQKTSKAAVPAGSTYNSLPDNDLAKLAVITGNAAATGNESVDSAAALAIQLD
LSGTGNDPTDISNIEGLQYLDNLSLNNISDLAPKLQVNLVSLNLSNRILV
NLGVEDLVNQLVNSANKALEDSQVSLPVLKEISAOQCNKITLEKNPAGAVLP
ELETFLQNDLNLTSKLPLKLVYKGNLSKLETLNGATKLQIDASNCITD
ETLDIGLSELEMQLSGSKLKEITSLKNLNLVNIADSCAIEDLGTNLLENK
TLVSDNENLTNITDLPQKLTLDGGCITSIGTLDNLPLEKLDLKENQITSI
EITDPLSLVDSVNNLTITDGLKPLLELWVNSNRSLSDVSLTNFPSLNYNIS
NNVIRTVKSNRLESTKFEYAQNNSISDI SMIHDPNLRKVDASNNLTINIGFDNLP
KLQSLDVHNSRTSLSTVHDLPSLETFAOQTNLTIGTMDNLPDLYVNLNFRNIPS
LAPIGDPLNLTIVSDNNSVLSRISGTMDGVKRLIIDLQNNYLVNTGTGECNLSLSD
LTNLTSLNLRNVVIDISGLTSLRLIYLNLSNKIEDISALSNLTNLOELTLNNK
IENISLSDLENLNLVSVKNNIIDSFVANNVRGAIVTASQTYTLPTVLSYQSSP
TIDNPVIAWYDGLTLPASSIGNSGNYKDKITWTNNATSSSLTFNFRNLKDLTFSGT
VTPYKAAKVTADEQTYTIGDTISEEQFKOVNAKSSDGPVTFATVNDITFG
EYEVTLTSEKDIGQSKVIVKLVHGAPVISADQTIISYDKHATITTEKQLEDIHAST
DLDTAITTNFSTAVLNKGGDYTVALNSENEDGVKAETVYTVTVNKKDPAIISAKTE

/transl_table=11
/protein_id="CAC98412.1"
/db_xref="GI:16409711"

/db_xref="SPTREMBL:O8YA32"

/translation="MKKFSIVISVLLGLVLPFDLLVCADETTVSEDVAVKTAEA
DSATEGLESTGDETAEPKEAEAEAKETTEREKAKTEEPASNIKTEINDKS
OLQKTSKAAVPAGSTYNSLPDNDLAKLAVITGNAAATGNESVDSAAALAIQLD
LSGTGNDPTDISNIEGLQYLDNLSLNNISDLAPKLQVNLVSLNLSNRILV
NLGVEDLVNQLVNSANKALEDSQVSLPVLKEISAOQCNKITLEKNPAGAVLP
ELETFLQNDLNLTSKLPLKLVYKGNLSKLETLNGATKLQIDASNCITD
ETLDIGLSELEMQLSGSKLKEITSLKNLNLVNIADSCAIEDLGTNLLENK
TLVSDNENLTNITDLPQKLTLDGGCITSIGTLDNLPLEKLDLKENQITSI
EITDPLSLVDSVNNLTITDGLKPLLELWVNSNRSLSDVSLTNFPSLNYNIS
NNVIRTVKSNRLESTKFEYAQNNSISDI SMIHDPNLRKVDASNNLTINIGFDNLP
KLQSLDVHNSRTSLSTVHDLPSLETFAOQTNLTIGTMDNLPDLYVNLNFRNIPS
LAPIGDPLNLTIVSDNNSVLSRISGTMDGVKRLIIDLQNNYLVNTGTGECNLSLSD
LTNLTSLNLRNVVIDISGLTSLRLIYLNLSNKIEDISALSNLTNLOELTLNNK
IENISLSDLENLNLVSVKNNIIDSFVANNVRGAIVTASQTYTLPTVLSYQSSP
TIDNPVIAWYDGLTLPASSIGNSGNYKDKITWTNNATSSSLTFNFRNLKDLTFSGT
VTPYKAAKVTADEQTYTIGDTISEEQFKOVNAKSSDGPVTFATVNDITFG
EYEVTLTSEKDIGQSKVIVKLVHGAPVISADQTIISYDKHATITTEKQLEDIHAST
DLDTAITTNFSTAVLNKGGDYTVALNSENEDGVKAETVYTVTVNKKDPAIISAKTE

ITYDKFSKTEAFLDDIDADTNDGSI VTSNFAFATVNLKAGDYTVTLNSINSDGVA
TPTAIIVHVEKIKIATISNTAQYKAKINETQFLKDVHASINASPITVAVLESDFE
TVKLDVPGTYTVTITATNEGVSAPKEVSVIRKIPAPEITADKEITYPKFEVSE
AEFLNDHATISDKNVALTSFSTVNLNKAAGDYTVTLNATNEDGVKATPVEIVHVQ
QSERPVITADATISVDKFNITAEKLEIDHATSDGQSSSTVITSNFQATNFKTAMS
YTVTLNATNEDGISAEPAVTVTINKPEAALKAADAEVYAKNEAVTESDFFKDVHLE
GTEATSKATQNFDSVDSKTDYTVTINATNEDGAVSTPIEVVHIEASAPVIT
ANAEVKNKHEQTDERRFLYDSEAKIDEANVEIKTDFAEKVDINKVGTIVTLTATNE
DCQANPEVSVIVSDAAAEKNVYVDENGSEISAAETLTGNLDETFFIDAKSIAGY
KCDATLSGVFSVEQTVVPHYKAIKPGVVTIKYEDTNGKAVDEKQITGEVGDDEAE
AQVSGISCRATASGKITEEPQTIITFTYSTATPSKKSGEITVQYVDESCKLADSKV
TGNISDTSYEAKEALGEGYVGDGSAKVGFTEKSTQVTFPKMNTQVSDKDDPKGT
NQSSTDTTKLVKNNSLPATGDTENMILAVLIGFNNLIVASIFLFRPKPTNQ"

gene

6491..7032
/gene="lmo0334"

RBS

6491..6496
/gene="lmo0334"

CDS

6508..7032
/gene="lmo0334"

/codon_start=1
/transl_table=11
/protein_id="CAC98413.1"
/db_xref="GI:16409712"

/db_xref="SPTREMBL:O8YA31"

/translation="MNKIIYSQYTSQYATELTSQKKAIPMYSVESIPESIAEAE
IIVGSIYMGKVLGPKFKSKYKTLSSKLIVISAGYDPTKENTDNIBLTVKENLQK
TNFLKDI FCLAGKLDTQQLRKLHKLIVNALYKSAKRKESDLNDRDVRAYEDDS
QIDLKLDGVDRL"

gene

7235..7240,7248..7532)
/gene="lmo0335"

RBS

7235..7240
/gene="lmo0335"

CDS

7248..7532
/gene="lmo0335"

/codon_start=1
/transl_table=11
/protein_id="CAC98414.1"
/db_xref="GI:16409713"

/db_xref="SPTREMBL:O8YA30"

/translation="KREIKVDERTFOOHATKLASESTGSLPLKNGNMAYSKANSDIQ
LRSALIELVDVVDFOHVTTKNSRLKRGAIYAKODQLMGKINOLEVR"

gene

7519..7524
/gene="lmo0336"

RBS

7519..7524
/gene="lmo0336"

gene

7533..7901
/gene="lmo0336"

CDS

7533..7901
/gene="lmo0336"

/codon_start=1
/transl_table=11
/protein_id="CAC98415.1"
/db_xref="GI:16409714"

/db_xref="SPTREMBL:O8YA29"

/translation="MDSAHSOLEQQLQKIKAKITAEADQTRRKQNEQDWLEEDSN
QLTQEKRVLLDFLRSGMQCEBASSPHRYLEEQQHESQVWRDLQDKRSLDTELOEN
KORLHLEAKQATLQEWSSQ"

gene

7885..7890
/gene="lmo0337"

RBS

7885..7890
/gene="lmo0337"

gene

7898..9373
/gene="lmo0337"

CDS

7898..9373
/gene="lmo0337"

/codon_start=1
/transl_table=11
/protein_id="CAC98416.1"
/db_xref="GI:16409715"

/db_xref="SPTREMBL:O8YA28"

/translation="MSRIDIAELNDFLHGLRSSNAEKAMIRKIKEAMMDYADNRK
GEAVSTKQFFSSTYISICSTIEALDESEERLAQIRFGSQVSDSPSPARIDAEILQ
EMAKVSLQRKEEDLHRQLTAPNTKPDQOVYAVKRSVHTQLLKAIBOENLEKYL


```
AFQSHGQFPFSADELIRATARAVALQELHHVSNFKTGTYSVPKSAANSLLMKKARD
KARKENDQPKPAFEEDRYTVLAYTYVNDQGETVYMWLEKDKRKRVENKLODFLEKHG
QELDPPLYTNLSSEEREDYKVNDSWKEGINLNGQKVSQVSGVATLRSSAVYASMKOWTD
DAGLTDMALGLGFCIAAARKNPVIKEKLNPAETKQTVNGEKWNNYFKEKYGACNVOWK
PSSFEDIILNPERLYGSTKEVPQMLGEGWKEEYTGSAAGNWKFTNEGDGMVPHYHCE
GIHKSYGYFPSSDGTQVKLVGKDYIDFSKDKATIIIFKFGS"
gene      9362..9368
/genes="lmo0338"
RBS      9362..9368
/genes="lmo0338"
gene      9377..9757
/genes="lmo0338"
CDS      9377..9757
/codon_start=1
/transal_table=11
/protein_id="CAC98417.1"
/db_xref="GI:16409716"
/db_xref="SPTREMBL:O8YA27"
/translations="MKKLTVGGKFPDIYDITLKSGMFILELSNEEIEITFFEDPIVGV
TSFSSKNLSSELKANETIIDDEMEKNAYLLREKVLKIDNTDLWNINSVRQSSSEWLDIFR
RDELAKROLKERSDEIEVYKTL"
gene      join(10019..10024,10032..10403)
/genes="lmo0339"
RBS      10019..10024
/genes="lmo0339"
CDS      10032..10403
/genes="lmo0339"
/notes="weakly similar to inorganic pyrophosphatase"
/codon_start=1
/transal_table=11
/protein_id="CAC98418.1"
/db_xref="GI:16409717"
/db_xref="SPTREMBL:O8YA26"
/translations="MWSRQSIQKQLSKTKQTLNWTIDRPIGYTDKFGNVYPLNYG
FVGGVGGGSEQDVYVLSRKISEPIDTFQGLDIAVITRNDVEDKWWVAAPNENFTI
EIREKVFIEQFVDSIEIKLI"
gene      join(10402..10407,10416..10634)
/genes="lmo0340"
RBS      10402..10407
/genes="lmo0340"
CDS      10416..10634
/genes="lmo0340"
/codon_start=1

Query Match      6.5%; Score 127.8; DB 1; Length 280050;
Best Local Similarity 51.3%; Pred. No. 4.4e-18;
Matches 325; Conservative 0; Mismatches 302; Indels 6; Gaps 1;

QY      96      TTGATTACGGTGTGGAAATGTTAGGAGTGTGAGGAACGCCATCCGCACACTTTGGATTG 155
Db      243522  TTGACTATGATACAGGAAACACGAGNAGCATCAGCAAGCACTTGATTTATCGACTAC 243463

QY      156      ATATCAAGAATGTGCAAAAGCCAGAGATTTCTAAATGCTAAGCGCCTTATCTTCTCCTG 215
Db      243462  AAAACAAAATTTCTAGTGATGCAACAGAAATCTCACAGGCTGACGGGTTATTTTACCAG 243403

QY      216      GCCTTGGGGCTTTGCACCTGCATGATGTTCTTATTTCGTAAGGCTGGCTGAAGCAC 275
Db      243402  GAGTTGGCGCTTATCTCTGAAGCTATGAAGAACTCACTCGCGCGGGCTAAATAAACAAT 243343

QY      276      TCTGTACTTACATTCAGAATGATCGACCTTCTCTGGGTATATGCTGGGATTTCGAGCTAC 335
Db      243342  TAAAGAAATCGTCCGGTGGCAACCAATTTCTGTGTGTTCTGTGTATGCACTTT 243283

QY      336      TCTTTGAGTCAAGTGAAGAAATATGGTCCAAATCAAGTCTTGGCTTGATTCCTGGACGGG 395
Db      243282  TACTCGAATCAAGTAACGACATAGCTTTTACTAATGCTCGGTCTCTATTCCCGGTCAATG 243223

QY      396      TTGGCGCTTTTGATCATCCAAATGGTTTAAGGGTGCCACATATTTGATGTCATGCCCTTGG 455
Db      243222  TCGAAAAATTCACGAAGATCCCGAATTTGCCGTTCCACATATGGGCTGGATCAATCAACTAC 243163
```

```
QY      456      ATATAAAGGAAGGTCAGCAATTTTAGATGATGTGGGGAATCAACATGTGATTTTGTTC 515
Db      243162  AAATCAACGTCGCCACCCCGCTTACAAAAACAACCTTGATGGAGAATACGTTATTATCGTT 243103

QY      516      ACTCATATCAGCCATGCGGAGGACAAACAAGAGTGGATTTTCATCTACATCAGCTATG 575
Db      243102  ACTCTTATTACGCCAATTCGCCAG-----AAGAATACATTAATCGCAACAAGTGGCTATT 243049

QY      576      GTGACGATTTTATGTCATCCATTCAGAGGGAAATGTTTCATGCAGTCCCAATTTTCATCCCG 635
Db      243048  CTATTGAAGTTCTTGGCATGATAATAAAGGCNAATATTACGGCGCCAGTTCCATCTCTG 242989

QY      636      AGAAGAGTGGAGTGTGGACTTTCCATATTGAGAAGATTTTGAATGCTGATTCCTTTA 695
Db      242988  AAAAAAGCGCCAAATCGGCTTTGAAATTTTAAAGGATTTAAGGAGGTCACTTATTCAT 242929

QY      696      ACAACAAGACAGAGCCCAATGAATGGAAGG 728
Db      242928  GCAATCTTTCCCGAGCAATAGATTTTAAAAAATGG 242896

RESULT 15
AX641665/C      349980 bp      DNA      linear      PAT 21-FEB-2003
LOCUS      Sequence 2855 from Patent WO0101118.
DEFINITION      AX641665
ACCESSION      AX641665
VERSION      AX641665.1 GI:28474426
KEYWORDS
SOURCE      Listeria monocytogenes
ORGANISM      Listeria monocytogenes
REFERENCE      Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
AUTHORS      Buchrieser, C., Frangeul, L., Couve, E., Rueniok, C., Fsihi, H.,
Dehoux, P., Dussart, O., Chetouani, F., Nedjari, H., Glaeser, P.,
Kunst, F., Cossart, P., Daniels, J., Goebel, W., Kreft, J., Kuhn, M.,
Ng, E., Vazquez-Boland, J., Dominguez-Bernal, G., Garrido-Garcia, P.,
Tierrez-Martinez, A., Amend, A., Chakraborty, T., Domann, E., Hain, T.,
Berche, P., Charbit, A., Durant, L., Perez-Diaz, J. C., Baquero, F.,
Garcia del Portillo, F., Gomez-Lopez, N., Madueno, E., de Pablo, B.,
Wehland, J., Kaerst, U., Entian, K. D., Hauf, J., Rose, M. and Voss, H.
TITLE      Listeria monocytogenes genome, polypeptides and uses
JOURNAL      Patent: WO 010118-A 2855 11-APR-2001;
INSTITUT PASTEUR (FR)
FEATURES
Location/Qualifiers
1..349980
/organism="Listeria monocytogenes"
/mol_type="genomic DNA"
/db_xref="taxon:1639"
/note="Seq 1 for big: original length: 2944528-splitted
into 10 <223>seq 0001 : From 0.000.001 to 0.349.980
length: 349.980 <223>seq 2855 : From 0.300.001 to
0.649.980 length: 349.980 <223>seq 2856 : From 0.600.001
to 0.949.980 length: 349.980 <223>seq 2857 : From
0.900.001 to 1.249.980 length: 349.980 <223>seq 2858 :
From 1.200.001 to 1.549.980 length: 349.980 <223>seq 2859
: From 1.500.001 to 1.849.980 length: 349.980 <223>seq
2860 : From 1.800.001 to 2.149.980 length: 349.980
<223>seq 2861 : From 2.100.001 to 2.449.980 length:
349.980 <223>seq 2862 : From 2.400.001 to 2.749.980
length: 349.980 <223>seq 2863 : From 2.700.001 to
2.944.528 length: 244.528"
BASE COUNT      115068 a 60319 c 71179 g 103414 t
ORIGIN
```

Query Match 6.5%; Score 127.8; DB 6; Length 349980;
Best Local Similarity 51.3%; Pred. No. 4.4e-18;
Matches 325; Conservative 0; Mismatches 302; Indels 6; Gaps 1;

```
QY      96      TTGATTACGGTGTGGAAATGTTAGGAGTGTGAGGAACGCCATCCGCACACTTGGATTG 155
Db      303522  TTGACTATGATACAGGAAACACGAGAGCATCAGCAAGCACTTGATTTTATCGGACTAC 303463

QY      156      ATATCAAGAATGTGCAAAAGCCAGAGATTTCTAAATGCTAAGCGCCTTATCTTCTCCTG 215
```

```
Db 303462 AAAACAAATTTCTAGTGATGCAACAGAAATCTCACAGGCTGACGGGTTATTTTACCAG 303403
Qy 216 GCGTTGGGGCTTTGCACTGCAATGATGTTCTTATTCGTAAAGGCTGGCTGAAGCAC 275
Db 303402 GAGTTGGCGCTTATCCTGAAGCTATGAAGAACTCACTCGGCGGGCTAAATAAACAT 303343
Qy 276 TCTGTACTTACATTCAGNATGATCGACCTTTCCTGGGTATATGCTGGGATTCAGCTAC 335
Db 303342 TAAAGAAATCGCTGCGCGGTGGCAACCAATTCCTGGTGTTCCTGGGTATGCACTTT 303283
Qy 336 TCCTTGAGTCAAGTGAAGAAATCGTCCAATTCAGGTCTTGGCTTCGATTCCTGGACGG 395
Db 303282 TACTCGAATCAAGTAAAGCAATAGCTTTACTAATGCTCGGTCTCATTTCCCGTATG 303223
Qy 396 TTGGCGCTTTTGAATCATCCAAATCGTTTAAAGGGTGCCACATATTTGGATGGATGCGCTGG 455
Db 303222 TCGAAAAATTACCAGAGATCCCGAATTTGCCGTTCCACATATGGGCTGGAATCAACTAC 303163
Qy 456 ATATAAGGAGGCTCAGCAATTTTAGATGATGTGGGGAATCAACATGCTATTTGTTTC 515
Db 303162 AAATCNAACGTGCCACCCCGCTTACAAACAACTTGATGGAGAAATACGTTTATACGTT 303103
Qy 516 ACTCATATCGAGCCAATGCCGAGGACAAACAAAGAGTGGATTTCATCTACATGCGACTATG 575
Db 303102 ACTCCTATTAGCCCAATTTGCCAG-----AGAAATACATTTATCGCAACAGTGGCTATT 303049
Qy 576 GTGACGATTTTATTGCAATCCATTCAGAAAGGAAATGTTTCATGCAATCCAAATTTTCATCCCG 635
Db 303048 CTATTGAAGTTCCTGGCATGATAAATAAAGGCAATATTACGGCGCCAGTTCATCCTG 302989
Qy 636 AGAAGAGTGGAGGTGTTGGACTTTCATATTTGAGAAGATTTTGAATGCTGATTCCTTTA 695
Db 302988 AAAAAAGCGCCAAATCGGTCTTGAAATTTTAAAGGATTTAAGGAGGTCACTTATTCAT 302929
Qy 696 ACAACAAAAGACAGAAGCCAAATGAATGGAAGG 728
Db 302928 GCAATCTCCAGCAATAGATTTAAAAATGG 302896
```

Search completed: January 20, 2004, 23:49:58
Job time : 7204 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 18:26:05 ; Search time 535 Seconds
(without alignments)

9869.347 Million cell updates/sec

Title: US-09-831-233A-1

Perfect score: 1956

Sequence: 1 gcacgagaaggaactggcaa.....aaaaaaaaaaaaaaaaaaaaa 1956

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512.

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1956	100.0	1956	21	AAA27325 Impatiens balsamia
2	521.6	26.7	1084	21	AAA27327 Corn putative glut
3	266.6	13.6	671	21	AAA27331 Soybean putative g
C 4	127.8	6.5	2944528	24	ABA01041 Listeria monocytog
C 5	125.4	6.4	1163020	24	ABQ67197 Listeria innocua c
C 6	125.4	6.4	3011208	24	ABQ69245 Listeria monocytog
7	117.2	6.0	4584	24	ABQ70963 Fusarium venenatum
8	113.2	5.8	1860	21	AAF08042

9	104	5.3	178	21	AAA27330 Soybean putative g
10	99.2	5.1	476	25	AB255560 Aspergillus oryzae
C 11	98.6	5.0	1664976	19	AAV21209 Methanococcus jann
12	91.2	4.7	1523	24	ABO69172 Listeria monocytog
13	91.2	4.7	2365589	24	ABA90521 Genomic sequence o
14	87.6	4.5	564	24	ABK77558 Bacillus clausii g
15	85.6	4.4	777	24	ABN92347 Staphylococcus epi
16	84.4	4.3	405	24	ABK77570 Bacillus clausii g
17	83.6	4.3	757	24	ABK73011 Bacillus lichenifo
18	83.6	4.3	757	24	ABK73180 Bacillus lichenifo
19	82.6	4.2	636	24	ABK73073 Bacillus lichenifo
20	80.8	4.1	744	22	AH52449 S. epidermidis ope
21	80.8	4.1	3022	22	AH54365 S. epidermidis gen
22	77.2	3.9	344	24	ABN77661 Human ORF2608 cDNA
23	75	3.8	6346	18	AAV74548 Staphylococcus aur
24	74.8	3.8	611	21	AAF08136 Fusarium venenatum
25	74.2	3.8	587	21	AAA27326 Corn putative glut
26	71.4	3.7	49646	21	AA81457 N. meningitidis pa
C 27	71.4	3.7	349980	21	AAF21608 Neisseria meningit
C 28	71.4	3.7	1437668	21	AAA81490 N. meningitidis B
C 29	69.8	3.6	494	23	ABV10021 Human prostate exp
30	69.8	3.6	495	21	ABA27329 Rice glutamine ami
31	68.4	3.5	774	22	AAH67256 C glutamicum codin
32	68.4	3.5	897	22	AAF71947 Corynebacterium gl
C 33	68.4	3.5	349980	22	AAH68531 C glutamicum codin
C 34	68.4	3.5	1830121	17	AA742063 Haemophilus influe
C 35	68.2	3.5	738	25	AB238917 N. gonorrhoeae nuc
36	67.8	3.5	765	25	AB238918 N. gonorrhoeae nuc
37	66.8	3.4	24417	18	AA797221 Pseudomonas aerugi
C 38	65.8	3.4	556	23	ABV40063 Human prostate exp
C 39	65.8	3.4	556	23	ABV40163 Human prostate exp
C 40	65.8	3.4	556	23	ABV42105 Human prostate exp
C 41	65.8	3.4	556	23	ABV43601 Human prostate exp
C 42	65.4	3.3	64081	24	AB92787 Buchnera sp. genom
C 43	65	3.3	309	23	ABV44994 Human prostate exp
C 44	64.2	3.3	419	25	ABX46069 Bovine EST associa
C 45	63.6	3.3	100848	22	AAF28552 Genomic fragment #

ALIGNMENTS

RESULT 1
AAA27325
ID AAA27325 standard; cDNA; 1956 BP.
XX
AC AAA27325;
XX
DT 08-AUG-2000 (first entry)
XX
DE Impatiens balsamia glutamine amidotransferase gene.
XX
KW Glutamine amidotransferase; histidine biosynthesis; herbicide;
KW fungicide; ss.
XX
OS Impatiens balsamia.
XX
FH Key Location/Qualifiers
FT CDS 2..1678
FT /tag= a
FT /product= "glutamine amidotransferase"
FT /partial
XX
PN WO200028053-A2.
XX
PD 18-MAY-2000.
XX
PF 04-NOV-1999; 99WO-US25950.
XX
PR 05-NOV-1998; 98US-0107275.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX

XX KW Glutamine amidotransferase; histidine biosynthesis; herbicide;
 XX KW fungicide; soybean; ss.
 XX OS Glycine max.
 XX FH Key Location/Qualifiers
 XX CDS 2..427
 XX FT /*tag= a
 XX FT /product= "glutamine amidotransferase"
 XX FT /partial
 XX FT /transl_except= (pos:275..277,aa:Xaa)
 XX FT /note= "Xaa=unknown"
 XX XX
 XX PN WO200028053-A2.
 XX XX
 XX PD 18-MAY-2000.
 XX XX
 XX PF 04-NOV-1999; 99WO-US25950.
 XX XX
 XX PR 05-NOV-1998; 98US-0107275.
 XX XX
 XX PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX XX
 XX PI Allen SM, Huang LL, Falco SC, Rafalski AJ;
 XX XX
 XX DR WPI; 2000-376564/32.
 XX DR P-PSDB; AAY94230.
 XX XX
 XX PT Novel polynucleotides encoding plant glutamine amidotransferase
 XX PT homologues, useful for producing transgenic plants and as probes or
 XX PT primers -
 XX XX
 XX PS Claim 2; Page 50; 53pp; English.
 XX XX
 XX CC The present sequence is a Glycine max (soybean) putative glutamine
 XX CC amidotransferase coding sequence. The protein forms a crucial stage in
 XX CC the histidine biosynthesis pathway. The sequence can be used to
 XX CC create transgenic plants which express different amounts of the protein,
 XX CC to identify loss of function mutants and to produce the protein in a host
 XX CC cell, for example a bacterium. The protein can be used to identify
 XX CC inhibitors which may be useful as fungicides and herbicides. The gene
 XX CC was identified by computer screening for sequences which could encode
 XX CC histidine biosynthetic enzymes.
 XX XX
 XX SQ Sequence 671 BP; 217 A; 104 C; 144 G; 205 T; 1 other;
 XX XX
 XX Query Match 13.6%; Score 266.6; DB 21; Length 671;
 XX Best Local Similarity 69.4%; Pred. No. 3.5e-50;
 XX Matches 376; Conservative 0; Mismatches 165; Indels 1; Gaps 1;
 XX
 Qy 1253 TACTTGAGAAACCCGATGAGTAGAATTTAAAGCCATCAAGTAAGCCATCCAGGTCCA 1312
 Db 2 TATGTAAGGATCTCTAATGATGTGCAATTTGAAGACCATTAAGGTTTCAAGTCCAGGTCCA 61
 Qy 1313 AACCGTCAGGATATGCTGTATCACTGCTTAAATGGTCAGCAGAGGAGGAGACCC 1372
 Db 62 AATGGAGGAAATATGATGTATCAATGTACAGTTAATGGGACGAGAGGCGCGCCA 121
 Qy 1373 ATCCGAGCTTATGAACCTAGCTTAAGCGTGTGGAGAACTTTGGAGCTGGAGAAATATTATTG 1432
 Db 122 ATTGGTCTTATGAACCTAGCAAAAGCAGTTGAAGAGCTTGGTGTGGTAAATACTACTT 181
 Qy 1433 AACTGCATATGATGTATGTCAGGAAAGGATTCGATATAGATTCGATCAAGCTAATA 1492
 Db 182 AATTGCATTTGATTCGACCGGTCAAGGAAAGGATTTGATGTAGATTAATTAAGTTGATA 241
 Qy 1493 TCCGATGCTGTACATCTCTTTATCGCAGCAGCGTCCAGAGTCGCTGATCACTTC 1552
 Db 242 TCAATGCTGTGAATATCCCTGTTATCGCAAGTANCGGTGCTGTCTCGTGAACACTTC 301
 Qy 1553 TCCGAAGTCTTTAATGAACCAACGCAATCTGCTGCCCTTGCAGCTGGCATTTTCCATCGC 1612

302 TCTGAGGTGTTCTATATAAAACAAATGTCATCAGCAGCACTTCTGCTGCAATTTTTCACAGG 361
 1613 AAAGAGGTTCCCAATTAAAGGCTGTTAAAGAGCACTTGTGAAGGAGGATTTGAAGTTAGA 1672
 362 AAAGAGGTGCTTATTCAGTCGGTAAAGAGCAATTTGTTGAAGGAGGATAGAAAGTTTCCA 421
 1673 TTGTGAAGGCGA-GAATCAGCTTGAAGAAAATTTTCATCTTGAAGTTCAATTTTGTACACAA 1731
 422 ATCTGATCATATGCAATTTAGTGGTAAATTTTGGAGATGCACTTCATAAATCAAACT 481
 1732 GAGATTTCTCTTCTTGGCCCTATGATATTTATTTATTTATTTGTTTCTATTTGAATT 1791
 482 TGTACGTTATCCCTTTTGTGTAATCTAAGCAAGCTCTCAAAATGTCAACCTTTTGGATC 541
 1792 AT 1793
 542 AT 543

RESULT 4
 ABA03041/c
 ID ABA03041 standard; DNA; 2944528 BP.
 XX AC ABA03041;
 XX XX
 XX DT 05-FEB-2002 (first entry)
 XX XX
 XX DE Listeria monocytogenes EGD-e genome sequence.
 XX KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 XX KW vitamin B12; bacterial infection; disease; ds.
 XX XX
 XX OS Listeria monocytogenes.
 XX XX
 XX PN WO200177335-A2.
 XX XX
 XX PD 18-OCT-2001.
 XX XX
 XX PF 11-APR-2001; 2001WO-FR01118.
 XX XX
 XX PR 11-APR-2000; 2000FR-0004629.
 XX XX
 XX PA (INSP) INST PASTEUR.
 XX XX
 XX PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
 XX PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
 XX PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
 XX PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 XX PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 XX PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 XX PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
 XX PI Rose M, Voss H;
 XX XX
 XX DR WPI; 2002-010914/01.
 XX XX
 XX PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 XX PT and prevention of Listeria and related bacterial infections, and
 XX PT related polypeptides -
 XX XX
 XX PS Claim 1; SEQ ID No 1; 192pp; French.
 XX XX
 XX CC The present sequence is the genome sequence of Listeria monocytogenes
 XX CC EGD-e. This sequence and fragments of this sequence are useful for
 XX CC selecting probes and primers for detecting genes in L. monocytogenes and
 XX CC related organisms, and to study genetic polymorphisms and other genomes.
 XX CC Proteins (AB047297-AB050149) expressed from the present sequence are
 XX CC useful for raising specific antibodies, identification of L.
 XX CC monocytogenes and related organisms, and for biosynthesis and
 XX CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and
 XX CC proteins encoded by it are also useful for selecting compounds that
 XX CC regulate gene expression and cell replication and modulate L.
 XX CC monocytogenes-related diseases. In addition, this sequence and proteins
 XX CC encoded by it are useful in pharmaceutical and vaccine compositions for

CC the treatment or prevention of infections by L. monocytogenes and related
CC organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;
Query Match 6.5%; Score 127.8; DB 24; Length 2944528;
Best Local Similarity 51.3%; Pred. No. 6.5e-18;
Matches 325; Conservative 0; Mismatches 302; Indels 6; Gaps 1;
QY 96 TTGATTACGGTGTGGAAATGTTAGGAGTGTGAGGAACGCCATCCGCACACTTGGATTG 155
DB 603522 TTGACTATGATCAGGAACACGAGAGCATCAGCAAGCACTTGATTTATCGGACTAC 603463
QY 156 ATATCAAGATGTCAAAGCCAGAGATATTTCTAATGCTAAGCGCTTATCTTCCCTG 215
DB 603462 AAAAAAATTTCTAGTGATGCAACAGAAATCTCACAGGCTGACGGCTTATTTTACCAG 603403
QY 216 GCGTTGGGGCTTTGACCTGCAATGATGTTCTTATTCTAAAGGCTGGCTGAAGCAC 275
DB 603402 GAGTTGGCGTTATCTGAGACTATGAAAGACTCTCGCGCGGGCTAAATAAACAAT 603343
QY 276 TCTGTACTTACATTCAGAATGATCGACCTTTCCTGGGTATATGCTGGGATTCGAGCTAC 335
DB 603342 TAAAGAAATCGCTGCGGGTGCAACCAATTTCTTGGTGTGTTGTTGATGCAACTTT 603283
QY 336 TCTTTGAGTCAAGTGAAGAAATGGTCCAAATTCAGTCTTGGCTTGAATTCCTGGAGGG 395
DB 603282 TACTCGAATCAAGTAAAGCAACATAGCTTTTACTAATGCTCTCGTCTCATTTCCCGGTCATG 603223
QY 396 TTGGGGCTTTTGATCATCAATGTTTAAAGGTTGCCACATATTTGGATGCGATGCTTGG 455
DB 603222 TCGAAAAATACAGAAAGATCCCGAATTTCCGGTTCCACATATGCGCTGGAATCAACTAC 603163
QY 456 ATATAAGGAAGGTCAGCAATTTTAGATGATGTGGGAATCAACATGTGTATTTTGTTC 515
DB 603162 AAATCAACGTCGCCACCCGCTTACAAAACAACCTTGATGGAGATACGTTTATACGTTTC 603103
QY 516 ACTCATATCAGGCAATCGCGAGACMAAAGAGTGGATTTCACTACATGAGCTATG 575
DB 603102 ACTCCTATTACGGCAATTTGCCAG-----AAGAAATACATTATCGCAACAAAGTGGCTATT 603049
QY 576 GTGACGATTTTATTGATCATTTCAGAGGGAATGTTTCATGCAGTCCCAATTTTCATCCCG 635
DB 603048 CTATTGAAGTTCCTGGCATGATAAATAAGGCAATATTTACGGCGCCAGTTCATCCTG 602989
QY 636 AGAAGATGGAGGTGTGGACTTTTCCATATTGAGAAGATTTTGAATGCTGATTCCTTTA 695
DB 602988 AAAAAAGCGCCAAATCGGCTTGAATTTTAAAGGAATTTAAGGAGTCACTTATTCAT 602929
QY 696 ACAACAAAGACAGAGCCCAATGAATGGAAG 728
DB 602928 GCAAACTTCCAGCAATAGATTAAAAAATGG 602896
RESULT 5
AB067197/c
ID AB067197 standard; DNA; 1163020 BP.
XX
AC AB067197;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria innocua contig DNA sequence #10.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX
OS Listeria innocua.
XX
PN WO200228891-A2.

XX 11-APR-2002.
XX 04-OCT-2001; 2001WO-FR03061.
XX 04-OCT-2000; 2000FR-0012697.
XX (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX Kunst F, Glaser P;
XX WPI; 2002-332479/37.
XX New genomic sequences from Listeria species, useful for detection,
XX treatment and prevention of infection, also related polypeptides,
XX antibodies and modulators -
XX Claim 5; SEQ ID 10; 180pp; French.
XX The present invention relates to nucleic acid sequences
XX (AB067188-AB071212) from Listeria sp. The sequences are useful as probes
XX and primers for identification and/or detection of Listeria (e.g. as
XX contaminants in foods, or mutational analysis) and for analysis of
XX gene expression. Proteins encoded by the nucleic acid sequences can be
XX used to screen for compounds that modulate gene expression, replication
XX and pathogenicity of Listeria (potential therapeutic agents), also for
XX treating infections by Listeria, and are useful as immunogens in
XX anti-Listeria vaccines.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1163020 BP; 388339 A; 197589 C; 235612 G; 341474 T; 6 other;
Query Match 6.4%; Score 125.4; DB 24; Length 1163020;
Best Local Similarity 51.4%; Pred. No. 1.7e-17;
Matches 319; Conservative 0; Mismatches 296; Indels 6; Gaps 1;
QY 63 CTGGTCAGCGGAGGATTCGTGGTGACCTTACTTACTTACGCTGCTGGAATGTTAGGA 122
DB 512326 CTACTAAAGAGGCTCTGTAATGATTTATTATTGATTTATGATACAGAAACACAGA 512267
QY 123 GTGTGAGGAACGCATCCGCACACTTGGATTTGATATCAAGATGTGCAAGACCCAGAGG 182
DB 512266 GCATCAGCAAGACACTTGATTTTATTTGCTCTCAAAATAAATTTCTAGCAACCGCG 512207
QY 183 ATATTCTAAATGCTAAGCGCTTATCTTTCTGGCGTTCGGGCTTTTGCACCTGCAATGG 242
DB 512206 AAAATTTGCAAGCGGATGCGTTATTTTACCTGGAGTCGGAGCTTTCCAGAACGATGC 512147
QY 243 ATGTTCTTATTCTGTAAGGCTGGCTGAAGCACTCTGTACTTACATTTCAGATGATCGAC 302
DB 512146 AGGAACTAGGTCGCGCGCTGCGATGTGACGTAAAGAAATGCGCAGAAATCTGGCAAC 512087
QY 303 CTTTCTCGGTATATGCTCGGATTCAGCTACTCTTTTCAGTCAAGTGAAGAAATGGTC 362
DB 512086 CGATGCTCGGTGTTGTCTTGATGCAACTGCTACTTGAATCAAGTGAAGAACATCTCT 512027
QY 363 CAATTCAAGGCTTGGCTTTGATTCTCTGGACGGGTTGGCGTTTGTGAATCATCAATGGTT 422
DB 512026 TTACTAAAGGACTCGGCTCTCATTCAGGACATGTGAAAAAATCTACCAGACGAACCTGGAT 511967
QY 423 TAAAGGTGCCACATATTGGATGCGATGCTTCGGATATAAGAGGCTGAGCAATTTTAG 482
DB 511966 TTGCGGTTCCGACATGGGCTGGAATCAACTTTGAAATAAAGACGACACCCACTCATA 511907
QY 483 ATGATGTGGGAATCAACATGTGTTTGTTCACCTCATATCGAGCAATGCGAGGACA 542
DB 511906 AAAAATCGGCGTGAATGATGCTATTACGTCATTCCTA-----CTATGAAATTTGTC 511853
QY 543 ACAAGAGTGGATTTTCATCTACATGCAGCTATGGTGAACGATTTTATTTGATCATTCAT 602

```
Db 511852 CAAACGAATACATTATTGCAACTAGCGGTTATTCTGTTGAAGTTCCAGATGATGATAATA 511793
Qy 603 AGGGAATGTTTCATGAGTCAATTTTCATCCGAGAGAGTGGAGGTGTTGGACTTTCCA 662
Db 511792 AAGCAATATTACGGGCCAGTTTCCACCAGAAAAAGCGGTCAAAATCGGGCTTGAAA 511733
Qy 663 TATTGAGAAGATTTTTCGAATG 683
Db 511732 TTTTAAAGGGTTTAAAGAGG 511712

RESULT 6
ABQ69245/c
ID ABQ69245 standard; DNA; 3011208 BP.
XX AC
XX AC
XX AC
XX 29-AUG-2002 (first entry)
XX DE
XX Listeria innocua DNA sequence #684.
XX KW
XX Antibacterial; Listeria; food contamination; mutational analysis;
XX KW
XX infection; ds.
XX OS
XX Listeria innocua.
XX FN
XX WO200228891-A2.
XX PD
XX 11-APR-2002.
XX PF
XX 04-OCT-2001; 2001WO-FR03061.
XX PR
XX 04-OCT-2000; 2000FR-0012697.
XX PA
XX (INSP ) INST PASTEUR.
XX PA
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX PI
XX Kunst F, Glaser P;
XX DR
XX WPI; 2002-332479/37.
XX DR
XX New genomic sequences from Listeria species, useful for detection,
XX PT
XX treatment and prevention of infection, also related polypeptides,
XX PT
XX antibodies and modulators
XX PS
XX Claim 5; SEQ ID 2058; 180pp; French.

The present invention relates to nucleic acid sequences
(CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
(CC and primers for identification and/or detection of Listeria (e.g. as
(CC contaminants in foods, or mutational analysis) and for analysis of
(CC gene expression. Proteins encoded by the nucleic acid sequences can be
(CC used to screen for compounds that modulate gene expression, replication
(CC and pathogenicity of Listeria (potential therapeutic agents), also for
(CC treating infections by Listeria, and are useful as immunogens in
(CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX SQ
XX Sequence 3011208 BP; 941651 A; 568176 C; 559189 G; 942192 T; 0 other;

Query Match 6.4%; Score 125.4; DB 24; Length 3011208;
Best Local Similarity 51.4%; Pred. No. 2.3e-17;
Matches 319; Conservative 0; Mismatches 296; Indels 6; Gaps 1;

Qy 63 CTGGTGCGAGCGGAGATTCTGTGTGACTTTTACTTGTATTCGGTGCTGGAAATGTTAGGA 122
Db 602591 CTACTAAGAGAGTCCCTGTAAATGATTGTTATATTGATTATGATACAGGAACACGAGA 602532
Qy 123 GTGTGAGGAGCGCCATCCGACACTTGGATTGTATCAAGATGTGCAAAAGCCAGGAGG 182
Db 602531 GCATCAGCAAGCACTTGATTTTATTTGCTCTTCAAAATAAAATTTCTAGCAACCCAGCGG 602472
```

RESULT 7

```
ABQ70963
ID ABQ70963 standard; DNA; 4584 BP.
XX AC
XX AC
XX AC
XX 29-AUG-2002 (first entry)
XX DT
XX DE
XX DE
XX Listeria monocytogenes 4b contig DNA sequence #905.
XX KW
XX Antibacterial; Listeria; food contamination; mutational analysis;
XX KW
XX infection; ds.
XX OS
XX Listeria monocytogenes 4b.
XX PN
XX WO200228891-A2.
XX XX
XX 11-APR-2002.
XX PF
XX 04-OCT-2001; 2001WO-FR03061.
XX PR
XX 04-OCT-2000; 2000FR-0012697.
XX PA
XX (INSP ) INST PASTEUR.
XX PA
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX PI
XX Kunst F, Glaser P;
XX DR
XX WPI; 2002-332479/37.
XX DR
XX New genomic sequences from Listeria species, useful for detection,
XX PT
XX treatment and prevention of infection, also related polypeptides,
XX PT
XX antibodies and modulators
XX PS
XX Claim 14; SEQ ID 3776; 180pp; French.
```

XX The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ71212) from *Listeria* sp. The sequences are useful as probes
CC and primers for identification and/or detection of *Listeria* (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
CC treating infections by *Listeria*, and are useful as immunogens in
CC anti-*Listeria* vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4584 BP; 1452 A; 1000 C; 914 G; 1212 T; 6 other;

Query Match 6.0%; Score 117.2; DB 24; Length 4584;
Best Local Similarity 51.0%; Pred. No. 2.2e-16;
Matches 306; Conservative 0; Mismatches 288; Indels 6; Gaps 1;
QY 84 TGGTGACTTTACTTGTGATTACGCTGCTGCAAAATTTAGGAGTGTGAGGAAGCCCATCGCA 143
DB 1544 TGAATTGTATTTAGTACGATACAGGAATACGAAACATCAGTAAGCCCTTGATT 1603
QY 144 CACTTGGATTGTATCAAAAGATGTGCAAAAGCCAGAGGATATTTCTAAATGCTAAGCGCC 203
DB 1604 TTATCGGGTTTACAAAACAAAATTTCTAGTATAAAACAGAAATCGCTCAAGCTGATGGTG 1663
QY 204 TTATCTTCTCGGTTGGGCTTTCACCTGCAATGGATGTTCTTATTCGTAAGGGC 263
DB 1664 TTATTTTGGCCGAGTGTGCTTATCCAGAGCCATGCAAGAACTCACTCGACGGGAT 1723
QY 264 TGGCTGAAGCACTCTGTACTTACATTTCAGAAATGATCGACCTTTCCTGGGTATATGCTGG 323
DB 1724 TAGATAAAACATTAAGAAATTTGCTACCCCTGTGTAACCAATTTCTGGGTTTGCTTG 1783
QY 324 GATTGCACTACTCTTTGAGTCAAGTGAAGAAATGCTCCAAATTCAGAGTCTTTGGCTGA 383
DB 1784 GTATGCAACTATTACTCGAATCAAGTAACGAACATAGCTATCTAAAGGGCTCGGCTCA 1843
QY 384 TTCTGACAGGGTGGCGTTTGAATCATCCAAATGTTTAAAGGTGCGCACATATTTGGAT 443
DB 1844 TTCCGGTCTGATCGAAATGCTACCAAGCAATCCGAAATTTGCCGTTTCCGCACATGGCT 1903
QY 444 GGCATGCTTGGATATAAGAGGGTCAACAATTTTAGATGATGTGGGGAATCAACATG 503
DB 1904 GGAATCAATTAACAATCAAGCAACCCCGCTTACAAACATCGCTGGAGATATG 1963
QY 504 TGTATTTTGTCTACTCATATCGAGCCAAATGCCAGGACACAAAGAGTGGATTTCATCTA 563
DB 1964 TCTATTACGTTTCATCTCTA-----CTATGCCAAGTGTCCAGAGCTTACATATCGCA 2017
QY 564 CATGCACTATGTGACGATTTTATTCATCCATTCAGAAAGGGAATGTTTCATGCGATCC 623
DB 2018 CAAGTGGATATGCTATCGATATCCCTGGCATGATAAATAACGGACATATTTACGGAGCAC 2077
QY 624 AATTTCATCCGAGAGAGTGGAGTGTTCGACTTTCATATTCAGAGATTTTGAATG 683
DB 2078 AGTTGCATCTCTGAAAAGAGCGGCAAAATCGGTCTTGAATTTTAAAGGTTTAAAGAGG 2137

RESULT 8

AAF08042
ID AAF08042 standard; cDNA; 1860 BP.
XX
AC AAF08042;
XX
DT 13-MAR-2001 (first entry)
XX
XX Fusarium venenatum EST SEQ ID NO:565.
DE Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW

KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
XX Fusarium venenatum.
XX
XX WO200056762-A2.
XX
XX 28-SEP-2000.
XX
XX 22-MAR-2000; 2000WO-US07781.
XX
XX 22-MAR-1999; 99US-0273623.
XX
XX (NOVO) NOVO NORDISK BIOTECH INC.
XX (NOVO) NOVO NORDISK AS.
XX
XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
XX WPI; 2000-594572/56.
XX
XX Monitoring differential expression of genes in filamentous fungal cells
XX uses fluorescence-labeled nucleic acids isolated from the cells and a
XX substrate of expressed sequence tags -
XX
XX Claim 86; Page 601-602; 3161pp; English.
XX
XX The present invention describes a method for monitoring differential
XX expression of genes in a first filamentous fungal (FF) cell relative to
XX expression of the same genes in one or more second filamentous fungal
XX cells. The method uses fluorescence-labeled nucleic acids isolated from
XX the FF cells and a substrate of expressed sequence tags (EST). The ESTs
XX are used in the methods for monitoring differential expression of genes
XX in a first filamentous fungal (FF) cell relative to expression of the
XX same genes in one or more second filamentous fungal cells. Monitoring
XX the global expression of genes from FF cells allows the production
XX potential of the microorganisms to be improved. New genes may be
XX discovered, possible functions of unknown open reading frames can be
XX identified and gene copy number variation and stability can be
XX monitored. The expression of genes can be used to study how FF cells
XX adapt to changes in culture conditions, environmental stress, spore
XX morphogenesis, recombination, metabolic or catabolic pathway
XX engineering. Using ESTs provides several advantages over genomic or
XX random DNA clones including elimination of redundancy as one spot on an
XX array equals one gene or open reading frame, and organisation of the
XX microarrays based on function of the gene products to facilitate
XX analysis of the results. AAF07478 to AAF11247 represents ESTs from
XX Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
XX niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
XX AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
XX all specifically claimed in the present invention.
XX
SQ Sequence 1860 BP; 505 A; 416 C; 506 G; 422 T; 11 other;

Query Match 5.8%; Score 113.2; DB 21; Length 1860;
Best Local Similarity 52.1%; Pred. No. 1.3e-15;
Matches 244; Conservative 0; Mismatches 224; Indels 0; Gaps 0;
QY 1307 GGTCCAAACGGTGAGGAATATGCTGGTATCATGTGTCACATGTTAATGGTGACAGAGGG 1366
DB 2 GGTCTTAAGGGGAGGAGTACTGTTGGTNCNCCTGCACAAATCAAGGGTGGCCGGAAC 61
QY 1367 AGACCCATCGAGCTTATGAATAGCTAGGCTGTTGAGGAACTTCGAGTTCGAGAAATA 1426
DB 62 CGGGATATGGATGTAGTGTAGCTTGTGCAANCGTTCGAGCGATGGGAACCGGANNAGATT 121
QY 1427 TTATTGAATGCAATTCATTGTGTGTGTCAAGGAAAGGATTCGATATAGATCTGTATCAAG 1486
DB 122 CTCCTCACTGTGATTGACAAAGGACGACCAACAGTGGTTCGACTTCGAGCTCATCAAC 181
QY 1487 CTAATATCCGATGCTGTGAACATTCCTGTTATCGCAAGCAGCGGTGACAGAGTCGTGAT 1546
DB 182 CAGGTCAAGGGAGCTGTCAAGATCCCCTCATCGCTCAAGTGGTGTGCTGCAACCCGGCT 241

PR 04-OCT-2000; 2000FR-0012697.
XX (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX Kunst F, Glaser P;
XX WPI; 2002-332479/37.
XX New genomic sequences from *Listeria* species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators
XX Claim 14; SEQ ID 1985; 180pp; French.
XX The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ71212) from *Listeria* sp. The sequences are useful as probes
CC and primers for identification and/or detection of *Listeria* (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
CC treating infections by *Listeria*, and are useful as immunogens in
CC anti-*Listeria* vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1523 BP; 457 A; 312 C; 357 G; 397 T; 0 other;
Query Match 4.7%; Score 91.2; DB 24; Length 1523;
Best Local Similarity 55.3%; Pred. No. 1.1e-10; Indels 0; Gaps 0;
Matches 177; Conservative 0; Mismatches 143;
Qy 1352 GGTGGACGAGGAGGAGACCCATCGGAGCTTATGCACTAGCTAAGGCTGTGAGGAACCTT 1411
Db 1004 GCGGTCGAATGATACCGGACTTGATGCTATTGAATGGCTAAAAAAGCTGTACAACTT 1063
Qy 1412 GGAGCTGGAGAAATATTATTGAATGCAATGATGTTGATGTCAGGAAAGGATTTCAT 1471
Db 1064 GGTGCTGTGAAATCTTACTAACAAGTATGGACGCGATGGAACCAAAAAACGGTTACGAT 1123
Qy 1472 ATAGATCTGATCAAGCTAATATCGATGCTGTGAACATTCCTGTTATCGCAAGCAGCGGT 1531
Db 1124 ATCCGCTCACAAGCAATTTACGACGAGTTCCGTCAGTCATGCTTCAGGCGGT 1183
Qy 1532 GCAGGAGTCGTGATCACTTCTCCGAAGTCTTTAATGAACCAACGCACTGCTGCCCTT 1591
Db 1184 TCGGGAATGCCGACATATGCGCGAGGTTTTTGAATAAACAAGGCAACGCGGCTCTC 1243
Qy 1592 GCAGCTGGCATTTTCCATCGCAAGAGGTTCCAAATTAAGGCTGTTAAGAGCAGCTTGTG 1651
Db 1244 GCTGCAAGCATTTTCCACTACGCTGCACTAAGCATCAAAAATGTCAAAAACAACCTTACTC 1303
Qy 1652 AAGGAGGATTTGAAGTTAG 1671
Db 1304 GAAAAAGGATGAATATTCG 1323
RESULT 13
ABA90521
ID ABA90521 standard; DNA; 2365589 BP.
XX ABA90521;
XX
DT 16-MAY-2002 (first entry)
XX Genomic sequence of *Lactococcus lactis* IL1403.
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese; ds.
XX *Lactococcus lactis* IL1403.
XX

PN FR2807446-A1.
XX 12-OCT-2001.
XX 11-APR-2000; 2000FR-0004630.
XX 11-APR-2000; 2000FR-0004630.
XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;
XX WPI; 2002-043418/06.
XX New nucleotide sequence useful in the identification of *Lactococcus*
PT lactis and related species -
XX Claim 1; SEQ ID 1; 2504pp; French.
XX The present invention is related to a *Lactococcus lactis* nucleotide
CC sequence (ABA90521) and related proteins (AB953300-AB955621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify *Lactococcus lactis* or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO200177334 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2365589 BP; 765914 A; 415261 C; 420487 G; 763927 T; 0 other;
Query Match 4.7%; Score 91.2; DB 24; Length 2365589;
Best Local Similarity 53.0%; Pred. No. 1e-09; Indels 0; Gaps 0;
Matches 195; Conservative 0; Mismatches 173;
Qy 1307 GGTCCAAACGGTCAGGAATATGCTGCTATCACTGCTGCTTAAATGGTGACGAGGAGGG 1366
Db 1236595 GCTAAAAACGACGACACCATAGGGGATATGATGCTATATCAAGGTGGCGCTGAAAT 1236654
Qy 1367 AGACCCATCGGAGCTTATGAACCTAGCTAAGGCTGTTGAGGAACCTTGGAGCTGGAGAAATA 1426
Db 1236655 GCAGGCTTTGACTTAGTTGATTCGGCTAAAAAGTGTGAAAGATTAGGAGCAGCGAAAT 1236714
Qy 1427 TTATTGAATGCAATTCATTTGATGTCAGGAAAGGATTCGATATAGTATCTGATCAAG 1486
Db 1236715 TTACTAACCTCAATGATTAAGGATGGAACAAACAGCAGGCTACGATTTGGAAATGCTGAAT 1236774
Qy 1487 CTAATATCCGATGCTGTGAACATTCCTGTTATCGCAAGCAGCGGTGCGAGTCTGCTGAT 1546
Db 1236775 GAGCTTTGACTGCTGTTAATATTCCTGTTGAGCAAGTGGCGGTTGTGGGAAGATTCG 1236834
Qy 1547 CACTTCTCCGAAGTCTTTAATGAACCAACGCACTCTGCTGCCCTTGCAGCTGGCAATTTTC 1606
Db 1236835 GACATTTGTAAGCTTTTCAAAATACAAGAAGCGATGCTGCCCTTGTGCTTATTATTT 1236894
Qy 1607 CATCGCAAGAGGTTCCAAATTAAGGCTGTTAAAGAGCAGCTTGTGTAAGGAGGAGATTGAA 1666
Db 1236895 CATTATGGGGAAGCAACACAGTTGATGAAGTTAAAGACGAACTTATAAAAAATAATATTTCCA 1236954
Qy 1667 GTTAGATT 1674
Db 1236955 GCAAGAA 1236962
RESULT 14
ABK77558
ID ABK77558 standard; DNA; 564 BP.
XX ABK77558;
XX
DT 13-AUG-2002 (first entry)


```
XX
SQ Sequence 777 BP; 269 A; 106 C; 173 G; 229 T; 0 other;

Query Match      4.4%; Score 85.6; DB 24; Length 777;
Best Local Similarity 53.2%; Pred. No. 1.6e-09;
Matches 181; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

Qy 1334 TATCAGTGCACCTGTTAATGGTGGACGAGAGAGGGAGACCCATCGGAGCTTATGACTAGCT 1393
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1394 AAGGCTGTTGAGGAACCTGGAGCTGGAGAAATATTATTGAACCTGATTCATTGATGGTGGT 1453
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1454 CAAGGAAAGGATTCGATATAGATCTGATCAAGCTAATATCCGATGCTGTGAACATTCCT 1513
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1514 GTTATCGCAGCAGCGGTGCGAGGAGTCGCTGATCAGTCTCCGAGTCTTTAATGAAACC 1573
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1574 AACGCATCTGCTGCCCTTGACGCTGGCATTTTCCATCGCAAGAGGTTCCAAATTAAGGCT 1633
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1634 GTTAAAGAGCATTGTTGAAGGAAGGATTGAAGTTAGAT 1673
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1673 ATTAAGATAAATGCGTGAAGGAGGTATCCTCGTGAGAT 775
```

Search completed: January 20, 2004, 21:50:04
Job time : 552 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 21:18:15 ; Search time 126 Seconds
(without alignments)
6851.948 Million cell updates/sec

Title: US-09-831-233A-1
Perfect score: 1956
Sequence: 1 gcacgaggaactggcaaa.....aaaaaaaaaaaaaaaaaaaaa 1956

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	98.6	5.0	1664976	4	US-08-916-421B-1
C 2	85.6	4.4	777	4	US-09-134-001C-1810
C 3	71.2	3.6	708	4	US-09-252-991A-12223
C 4	68.4	3.5	1830121	4	US-09-557-884-1
C 5	68.4	3.5	1830121	4	US-09-643-990A-1
C 6	66.8	3.4	24417	2	US-08-846-762-1
C 7	63.8	3.3	807	4	US-09-328-352-2061
C 8	63.8	3.3	882	4	US-09-252-991A-6779
C 9	63.8	3.3	936	4	US-09-252-991A-6956
C 10	59	3.0	297	4	US-09-313-294A-6451
C 11	55.4	2.8	1664976	4	US-08-916-421B-1
C 12	54.4	2.8	7218	1	US-08-232-463-14
C 13	53.2	2.7	1559	4	US-09-489-847-42
C 14	51.6	2.6	489	4	US-09-252-991A-12336
C 15	51.6	2.6	798	4	US-09-252-991A-12156
C 16	51	2.6	624	4	US-09-328-352-1928
C 17	50	2.6	473	1	US-08-764-100-16
C 18	50	2.6	4970	1	US-08-764-100-14
C 19	50	2.6	4970	1	US-08-764-100-20
C 20	49.8	2.5	240	1	US-08-628-417-6
C 21	49.8	2.5	6409	4	US-09-967-908A-1
C 22	49.6	2.5	636	3	US-08-998-416-1137
C 23	49.6	2.5	2447	2	US-09-014-969-14
C 24	48.4	2.5	19124	2	US-08-487-826B-13
C 25	48.2	2.5	10640	4	US-09-417-485D-5
C 26	47.4	2.4	1696	4	US-09-835-811-1
C 27	46.8	2.4	2695	4	US-09-706-197-3

C 28	46.6	2.4	1447	4	US-09-443-041A-27	Sequence 27, Appl
C 29	46.4	2.4	1641	1	US-08-300-903A-8	Sequence 8, Appl
C 30	46.4	2.4	1641	4	US-08-988-197-8	Sequence 8, Appl
C 31	46.2	2.4	664	4	US-09-904-615-66	Sequence 66, Appl
C 32	45.4	2.3	5852	1	US-07-867-106-2	Sequence 2, Appl
C 33	45.4	2.3	6243	2	US-09-056-073-1	Sequence 1, Appl
C 34	45.2	2.3	1129	4	US-09-227-357-40	Sequence 40, Appl
C 35	45.2	2.3	2058	2	US-08-749-391-1	Sequence 1, Appl
C 36	45.2	2.3	2058	3	US-09-390-200-1	Sequence 1, Appl
C 37	45	2.3	2773	4	US-09-996-243-178	Sequence 178, App
C 38	45	2.3	8920	2	US-08-446-855A-1	Sequence 1, Appl
C 39	45	2.3	8920	3	US-09-150-741-1	Sequence 1, Appl
C 40	44.8	2.3	1117	3	US-09-247-373B-33	Sequence 33, Appl
C 41	44.8	2.3	4403765	3	US-09-103-840A-2	Sequence 2, Appl
C 42	44.8	2.3	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C 43	44.6	2.3	6671	1	US-08-280-443-1	Sequence 1, Appl
C 44	44.6	2.3	6671	1	US-08-457-459-1	Sequence 1, Appl
C 45	44.6	2.3	6671	1	US-08-555-678-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus jannaschii
; Patent No. 6503729
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)

```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g
;
US-08-916-421B-1

Query Match      5.0%; Score 98.6; DB 4; Length 1664976;
Best local Similarity 51.0%; Pred. No. 9.8e-14;
Matches 233; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

Qy 1259 AGAAACCCGATGAAGTAAAGTAAAGCCATCAAGTAAAGCCATCCAGGTCCAAACGGT 1318
Db 1101055 AGACACTATGTTAATGAAGATGAGATAGATAAATAAATGTCGTTAAAGTAGAG 1100996
Qy 1319 GAGGAATATGCTGGTATCATGTCACCTGTTAATGGTGGACGAGAGGAGACCCATCGGA 1378
Db 1100995 GATGGTTATTGCTGTTTGAAGTTTATATATACGGAGGGAGAGAAACAGGTATAGAT 1100936
Qy 1379 GCTTATGAAGTACTAGTAAAGCTGTTGAGGACTTGGAGCTCGAGAAATATTATTGAAGTGC 1438
Db 1100935 GCCATAAATCGGGCTGAGAAATTTGAAGATTTGGAGCTGGAGAGATTTTATTGCAAGT 1100876
Qy 1439 ATTGATTGTGATGGTCAAGGAGAAAGGATTCGATATAGATCTGATCAAGCTAATATCCGAT 1498

```


GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match 3.5%; Score 68.4; DB 4; Length 1830121;
Best Local Similarity 52.9%; Pred. No. 2.8e-06;
Matches 147; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
Qy 1394 AAGGCTGTGAGAACTTGGAGCTGGAGAAATATTGAACTGCAATTCGATTCCT 1453
Db 496613 AAAGAAGTACAACTGCGTGCAGGAGAAATCGTCTAAATATGATGAATCAAGATGGT 496672
Qy 1454 CAAGGAAAGGATTCGATATAGATCTGATCAAGCTTAATATCCGATGCTGCAATTCCT 1513
Db 496673 TTACGAAACGGGTACGATTTAGCCAGCTCAAGCTGGTGGAGGATGATGAGATACCT 496732
Qy 1514 GTTATCGAAGCAGCGGTGAGAGTGCCTGATCACTTCTCCGAAGTCTTTAATGAAC 1573
Db 496733 CTAATTGATCTGCGGTGCGGTAAATGTAATTTCTGATGCTTTTATTGAAGCA 496792
Qy 1574 AACGCATCTGTCCTTCGAGCTGGCATTTTCCATCGCAAGAGGTTCCAAATTAAGGCT 1633
Db 496793 AAAGTAGATGGGATAGCTGCGAGGATTTCCATAACAGATTAATTGAATGGCGAG 496852
Qy 1634 GTTAAAGAGCATTGTTGAAGGAGGATTTGAAGTTAG 1671
Db 496853 TTAAATCTCTACCTTGTGCAATCAGCCATCGAGATTCG 496890

RESULT 5
US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Mark D. Adams

Owen White
Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 3.5%; Score 68.4; DB 4; Length 1830121;
Best Local Similarity 52.9%; Pred. No. 2.8e-06;
Matches 147; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
Qy 1394 AAGGCTGTGAGAACTTGGAGCTGGAGAAATATTGAACTGCAATTCGATTCCT 1453
Db 496613 AAAGAAGTACAACTGCGTGCAGGAGAAATCGTCTAAATATGATGAATCAAGATGGT 496672
Qy 1454 CAAGGAAAGGATTCGATATAGATCTGATCAAGCTTAATATCCGATGCTGGAACATTCCT 1513
Db 496673 TTACGAAACGGGTACGATTTAGCCAGCTCAAGCTGGTGGAGGATGATGAGATACCT 496732
Qy 1514 GTTATCGAAGCAGCGGTGAGAGTGCCTGATCACTTCTCCGAAGTCTTTAATGAAC 1573
Db 496733 CTAATTGATCTGCGGTGCGGTAAATGTAATTTCTGATGCTTTTATTGAAGCA 496792
Qy 1574 AACGCATCTGTCCTTCGAGCTGGCATTTTCCATCGCAAGAGGTTCCAAATTAAGGCT 1633
Db 496793 AAAGTAGATGGGATAGCTGCGAGGATTTCCATAACAGATTAATTGAATGGCGAG 496852
Qy 1634 GTTAAAGAGCATTGTTGAAGGAGGATTTGAAGTTAG 1671
Db 496853 TTAAATCTCTACCTTGTGCAATCAGCCATCGAGATTCG 496890

RESULT 6
US-08-846-762-1
Sequence 1, Application US/08846762A
Patent No. 5994072

; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; APPLICANT: de Kievit, Teresa
; TITLE OF INVENTION: No. 594072el Proteins Involved in the Synthesis and Assembly
; FILE REFERENCE: of O-Antigen in Pseudomonas Aeruginosa
; CURRENT APPLICATION NUMBER: 6580-089
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 24417
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-08-846-762-1

Query Match 3.4%; Score 66.8; DB 2; Length 24417;
Best Local Similarity 46.5%; Pred. No. 1.3e-06;
Matches 288; Conservative 0; Mismatches 327; Indels 5; Gaps 2;

QY 70 AGCGGAGATCTCTGGTACTTACTTGTATACGGTCTGGAATGTTAGGAGTGTGAG 129
Db |||||
QY 130 GAACGCCATCCGCACACTTGGATTTGATATCAAGATGTGCAAAAGCCAGAGATATCT 189
Db |||||
QY 9826 GAACATGCTGAAGGAGTTGGTGCAAAGCCAGGCATCCGATAGCGAGAGATATCGA 9885
Db |||||
QY 190 AATGCTTAAGCGCTTATCTTTCTCGGCTTGGGCGCTTTGCACTGCAATGGATGTTCT 249
Db |||||
QY 9886 GCAGCGGAGAACTGATTTTGGCTGTGCTGGTCTTTGACCGCGAATGCAAACT 9945
Db |||||
QY 250 TATTCTTAAGGGTGTGCTGAAGCACTCTGTACTTACATTCAGAAATGATGCA---CCTTT 306
Db |||||
QY 9946 ACACAAGAGTGGGCTGGTGGATGTACTGACAGACAGGTGATGATCAACAGAAAGCGGT 10005
Db |||||
QY 307 CTTGGGTATATGCTGGGATTTGAGCTACTCTTTGAGTCAAGTGAAGAAATGTTCCAAT 366
Db |||||
QY 10006 CATGGGGTGTCTCGGGAGTCAAGA--TGCTGGGCTCGGATCTGAGGAGGAGCGGA 10063
Db |||||
QY 367 TCAAGGTCTTGGCTTGATTCCTGACGGGTGGGCGTGGTGGTGGTGGTGGTGGTGGTGGT 426
Db |||||
QY 10064 ACCGGGCTGGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10123
Db |||||
QY 427 GGTGCCACATATGGGATGCGATGCTTGGATATAAGGAAGGGTCAAGCAATTTTATGATGA 486
Db |||||
QY 10124 GGTTCACATATGGGCTGGAATCAAGTGTCCCGCAATTTGGAGCATCTTACTACTTAGCGG 10183
Db |||||
QY 487 TGTGGGGAATCAACATGTTGTTTGTCTACTCATATCGAGCCAATGCCGAGGACACAA 546
Db |||||
QY 10184 TATAACGACGAAGCGGATCTTATTTGTTCTATAGTTATATATGTTCCGAAAGACCC 10243
Db |||||
QY 547 AGAGTGGATTTCTATCATGCGAGTATGTTGAGCAATTTTATGCAATTCATTCAGAGGG 606
Db |||||
QY 10244 AGAGTATCTGTTGAGTTGTAATATGACAAATTTCACTCGCGCGTGGCTCGGA 10303
Db |||||
QY 607 AATGTTTCAGTCCAAATTCATCCCGAGAGAGTGGAGGTGTTGGCACTTTCCATATT 666
Db |||||
QY 10304 TAATGTTTTCCGATTTTCAGTTTCATCTCGAAGAGTGCATAAATTCGGTATGCAATTAT 10363
Db |||||
QY 667 GAGAAGATTTTGAATGCTG 686
Db |||||
QY 10364 CAAAACTCTGTGAGCTTG 10383
Db |||||

RESULT 7
US-09-328-352-2061
; Sequence 2061, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2061
; LENGTH: 807
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2061

Query Match 3.3%; Score 63.8; DB 4; Length 807;
Best Local Similarity 49.5%; Pred. No. 1.9e-06;
Matches 164; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 1348 TAATGCTGACGAGAGGAGACCCATCGAGCTTATGAAGCTTAAAGCTGTTGAGGA 1407
Db |||||
QY 477 TCACGGTGGCGTAAACCAACCGGTATGATGCCATTGGAATGGCTGTAAAAATGCTGA 536
Db |||||
QY 1408 ACTTGAGCTGGAGAAATATATTGAATCTGCATTGATTGTGATGGTCAAGGAAAAAGATT 1467
Db |||||
QY 537 CTACGGTGCAGGTGAGTTACTCATACAGTATGAGCGCGGCTACTAAACGAGTTA 596
Db |||||
QY 1468 CGATATAGATCTGATCAAGCTAATATCCGATGCTGTGAACATTCCTGTTTATCGCAAGCAG 1527
Db |||||
QY 597 CGATATTGCGTTAATCGGTGCTATTAAATGACCGGTGACTATTTCCAAACGATCGCATCAGG 656
Db |||||
QY 1528 CGGTGACGAGTCCGTGATCATCTTCCGAAGTCTTTAATGAACCAACGATCTGCTGC 1587
Db |||||
QY 657 CGGTGTGAGTAACTTACAACATCTTGCAGACGGTATTTTACAAGGTGGTCCGATGCAGT 716
Db |||||
QY 1588 CCTTGACGCTGGCATTTTCCATCGCAAGAGGTTCCTCAATTAAGGCTGTTAAAGACGACTT 1647
Db |||||
QY 717 ATGGCAGCAAGTATTTTCCACTTTGGCCCAATATACGATTCGAAAGCTAAACAATATTT 776
Db |||||
QY 1648 GTTGAAGGAGGAGGATTGAAGTTAGATTGTAA 1678
Db |||||
QY 777 GCGTGAACAGGTATCGAAATGCGCTTTGTA 807
Db |||||

RESULT 8
US-09-252-991A-6779/c
; Sequence 6779, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6779
; LENGTH: 882
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6779

Query Match 3.3%; Score 63.8; DB 4; Length 882;
Best Local Similarity 51.6%; Pred. No. 1.9e-06;
Matches 146; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1391 GCTAAGGCTGTTGAGGAACCTCGAGCTGGAGAAATATTATTGAACCTGCATTGATTGTGAT 1450
Db |||||
QY 375 GCGAAGATGAAGACCTGGCGCTGGAGATTCTCTGACCAAGCATGACACAGGAC 316
Db |||||
QY 1451 GGTCAAGGAAAAAGGATTCGATATAGATCTGTGATCAAGCTAATATCCGATGCTGTGAACATT 1510
Db |||||

Db 315 GGCGTGAAGAGCGGTTACGACCTGGCGGTGACCGCGCCATCAGCGAGCGGTGAACGTG 256
Qy 1511 CCTTTATCGCAACAGCGGTGACGAGTGGCTGATCACTTCTCGAAGTCTTTAATGAA 1570
Db 255 CCGGTGATCGTTCCGGCGGCGTGGCAACCTGGAGCACCTGGCGCGCGGATCCTCGAG 196
Qy 1571 ACCAAGCATCTGCTGCGCTTGACAGCTGGCATTTTCCATCGCAAGAGGTTCCAAATTAAG 1630
Db 195 GGCAGGCGGAGCGGTGCTCGCGCGGAGCATCTTCCACTTGGCGGAGTACACCGTGCCG 136
Qy 1631 GCTTTAAAGACACTTTGTTGAAGAGGAGGATTAAGTTAGAT 1673
Db 135 GAAGCAAGGCGTACCTGGCGGCGGTATCGTGGTGGCT 93

RESULT 9

US-09-252-991A-6956
; Sequence 6956, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6956
; LENGTH: 936
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6956

Query Match 3.3%; Score 63.8; DB 4; Length 936;

Best Local Similarity 51.6%; Pred. No. 2e-06;
Matches 146; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
Qy 1391 GCTAAGGCTGTGAGGAACCTGGAGCTGGAGAAATATTATGAACTGATGATTTGAT 1450
Db 652 GCGAAGAGATGGAAGACCTGGCGCTGGCGAGATTCTCTGACCAAGTGGACCGAC 711
Qy 1451 GGTCAAGGAAAGATTGATATAGATCTGATCAAGCTATATCCGATGCTGTGAACATT 1510
Db 712 GCGGTGAAGAGCGGTTACGACCTGGCGGTGACCGCGCCATCAGCGAGGCGGTGAACGTG 771
Qy 1511 CCTGTTATCGCAAGCAGCGGTGACAGGAGTGGCTGATCACTTCTCCGAAGTCTTTAATGAA 1570
Db 772 CCGGTGATCGTTCCGGCGGCGTGGCAACCTGGAGCACCTGGCGCGGATCCTCGAG 831
Qy 1571 ACCAAGCATCTGCTGCGCTTGACAGCTGGCATTTTCCATCGCAAGAGGTTCCAAATTAAG 1630
Db 832 GGCAGGCGGAGCGGTGCTCGCGCGGAGCATCTTCCACTTGGCGGAGTACACCGTGCCG 891
Qy 1631 GCTTTAAAGACACTTTGTTGAAGAGGAGGATTAAGTTAGAT 1673
Db 892 GAAGCAAGGCGTACCTGGCGGCGGTATCGTGGTGGCT 934

RESULT 10

US-09-313-294A-6451
; Sequence 6451, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313.294A

; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 6451
; LENGTH: 297
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700351825H1
; NAME/KEY: unsure
; LOCATION: 101, 122
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-6451

Query Match 3.0%; Score 59; DB 4; Length 297;

Best Local Similarity 56.9%; Pred. No. 1.9e-05;
Matches 107; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
Qy 29 CCACAGAATGGGTTCAAGATTCCGCGCGCTTGGCTGGTGCAGGCGGAGATTCTGTGGTG 88
Db 110 CCGCGATGGGTGCGGTCTTCCGTCTCGTCCGTGCGTCTGCGCAACACGGTG 169
Qy 89 ACTTTACTTGATTACCGTGTCTGAAATGTTAGAGTGTGAGGAACGCCATCCGCACATT 148
Db 170 ACTCTGTGACTACGCGCGCGGAACGTACGCGCGTGCACCAACGCAATTTCGTACCTC 229
Qy 149 GGAATTTGATATCAAGATGTCGAAAGCCAGAGGATTTCTAAATGCTAAGCGCCTTATC 208
Db 230 GGCTTCGACATCCGCGACGTGCAGAGCCCGGAGGACATCGTCCCGCGGAACGGCTCGTT 289
Qy 209 TTTCTCTGG 216
Db 290 TTTCCCGG 297

RESULT 11

US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 6503729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature

/ LOCATION: (98120)..(98120)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (98159)..(98159)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (98239)..(98239)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (98266)..(98266)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (98343)..(98343)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (10398)..(10398)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (148948)..(148948)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (163385)..(163385)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (191989)..(191989)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (191995)..(191995)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (231980)..(231980)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (234187)..(234187)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (234220)..(234220)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (234814)..(234814)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (309398)..(309398)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (309418)..(309418)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (312837)..(312837)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (312993)..(312993)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (319226)..(319226)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (559167)..(559167)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (559241)..(559241)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (600992)..(600992)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (622708)..(622708)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (657081)..(657081)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (657203)..(657203)

/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (674435)..(674435)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (682442)..(682442)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (713652)..(713652)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (741684)..(741684)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (779455)..(779455)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (779676)..(779676)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (855539)..(855539)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (871619)..(871619)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1084830)..(1084830)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1096846)..(1096846)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1119881)..(1119881)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1130881)..(1130881)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1310988)..(1310988)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1313224)..(1313224)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1349473)..(1349473)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1349491)..(1349491)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1470091)..(1470091)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1569020)..(1569020)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1602912)..(1602912)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1603734)..(1603734)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1637998)..(1637998)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1664854)..(1664854)
/ OTHER INFORMATION: n equals a, t, c, or g
/ US-08-916-421B-1

Query Match 2.8%; Score 55.4; DB 4; Length 1664976;
Best Local Similarity 50.8%; Pred. No. 0.0044;
Matches 299; Conservative 0; Mismatches 266; Indels 24; Gaps 6;

QY 93 TACTTGTACCGTCTGGAAATGTTAGGAGTGTGAGGAACGCCATCCGACACTTGGAT 152
Db 1024660 TAATTGATTAACAGCGAGGATTTGAGAGTAT-TCAAAGGAGCTTGAATCTATGAT 1024718
QY 153 TTGATATCAAAAGATGTGCAAAAGCCAGAGATATCTAAATGCTAAAGCGCTTATCTTTC 212
Db 1024719 AAGGTAAATAACAAACACAG--TGAGGAGTATTGGCTGTGTAAGATAATCTTAC 1024776
QY 213 CTGCGTGGGGCTTTGACCTGCAATGATGTTCTTATTCGTAAGGCTGGCTGAAG 272
Db 1024777 CAGGTGTAGGAAATTTTGGTAGTGCATGGAATTTAGTCTC-----CATTAAGAGAGA 1024830
QY 273 CACTCTGTACTTACATGATCGACCTTTCTGGGTATATGCTCGGATTCGAGC 332
Db 1024831 CAATATCAAAATTTGATGATAGATGTTCCATCTTAGGAATATGTTTAGGAATGAGA 1024890
QY 333 TACTCTTTGAGTCAAGTGAAGAAATGCTCCAAATTCAGGTCTTTGGCTTGATTCCTGGAC 392
Db 1024891 TTTTATTTGAAGAGAGCGAAGAGAAAGAGGAATCAAAGGTTTAGGGATAATAAAGGCA 1024950
QY 393 GGGTGGGGCTTTGAATCATCCATGTTTAAGGGTCCACATATTTGGATGCGATGCCCT 452
Db 1024951 ATGTAATCAAGTTTAA-----GGATGTTCAAAACTTCCACATATGGGCTGGAATAGTG 1025004
QY 453 TGGATATAAAGGAGGCTCAGCAATTTTAGATGATGTGGGAATCAACAT---GTGTATT 509
Db 1025005 TAAATATGTTAAGATTGCCACTGTTTGAAGGATAAATAACATAGTACTTTTACT 1025064
QY 510 TTGTTCACTCATATCGAGCCAAATGCCGAGGACAAAGAGTGGATTTTCATCATGCA 569
Db 1025065 TTGTTCACTCATATCATGTAATCCAGATG-----AAGATTGTATAGTTGGAAGAACTG 1025118
QY 570 GCTATGTGAGGATTTTATGATCCATTCAGAGGGAATGTTTCAGTGCATTCATTTTC 629
Db 1025119 AATATGAAGAGAGTTTCCAGCGCTTATAAACAAGATAATGTTTTCGCCACCAATTC 1025178
QY 630 ATCCGAGAGAGTGGAGGTGTGGACTTTCCATATTGAGAAGATTTT 678
Db 1025179 ACCGAGAAAGTGGGAAATTTGTTTAAAGATTATAGAAAATTTTGT 1025227

RESULT 12

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCK NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZ9pt-Fls
; US-08-232-463-14

Query Match 2.8%; Score 54.4; DB 1; Length 7218;

Best Local Similarity 2.6%; Pred. No. 0.00091;
Matches 10; Conservative 224; Mismatches 150; Indels 0; Gaps 0;

QY 1086 AGTATTTTCAGATCGGCGCGGATAGAGTTTCGATCGGAAGTATGACAGTTTACACTGCTG 1145
Db 1446 AGAATTTGTACRR 1387
QY 1146 AGGAATATATTAACACCGGAGTGAGACAGAGAAAGACAGACATAGACAGATATCTACAG 1205
Db 1386 RRR 1327
QY 1206 TATATGTTACACAGGAGTGTTGTAAGCATTTGATCTCGCCGAGTTTACTTGAGAAAC 1265
Db 1326 RRR 1267
QY 1266 CCGATCAAGTAGAATTTAAAGCCATCAAAGTAAGCCATCCAGTCCAAACGGTGAAGAAT 1325
Db 1266 RRR 1207
QY 1326 ATSCCTGGTATCAGTCACCTGTTAAATGTTGAGAGAGAGAGACCCATCCGAGCTTATG 1385
Db 1206 RRR 1147
QY 1386 AACTAGTAAGCTGTTGAGGAATTTGAGACTGGAGAAATATTATTGAACCTGATTGATT 1445
Db 1146 RRR 1087
QY 1446 GTGATGTCAGAAAGGATTCG 1469
Db 1086 RRR 1063

RESULT 13

US-09-489-847-42/c
; Sequence 42, Application US/09489847
; Patent No. 6476195
; GENERAL INFORMATION:
; APPLICANT: Rosen et al
; TITLE OF INVENTION: 98 Human Secreted Proteins
; FILE REFERENCE: F2031PI
; CURRENT APPLICATION NUMBER: US/09/489,847
; CURRENT FILING DATE: 2000-01-24
; EARLIER APPLICATION NUMBER: PCT/US99/17130
; EARLIER FILING DATE: 1999-07-29
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 60/095,486
; EARLIER FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 60/096,319
; EARLIER FILING DATE: 1998-08-12
; EARLIER APPLICATION NUMBER: 60/095,454
; EARLIER FILING DATE: 1998-08-06
; EARLIER APPLICATION NUMBER: 60/095,455
; EARLIER FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 376

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; LENGTH: 1559
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-489-847-42

```

Query Match	2.7%	Score 53.2;	DB 4;	Length 1559;
Best Local Similarity	54.6%	Pred. No. 0.00098;		
Matches 106;	Conservative	0;	Mismatches 88;	Indels 0

Qy	1763	TTTTATTATTATGTTTGGCTATTGAAATTAATGTTATATATTTGGCAATTTGTTATTTT	1822
Db	1558	TTT	1499
Qy	1823	GAATAGATTTGAGTGTTTTAGACCTTGGTGTGCTCTGTTTATCTCTAGGCCCATGTTTGTG	1882
Db	1498	TTTTTGGCTTCTCTGAATAGATTTTATTGTTGTTTGGCTGATCTGAATTTGGAAATTTATA	1439
Qy	1883	GATTATATACAACTGTGAATTTAAATAAATTAATTCGATGAATTTATGCTTTTAAAAAAA	1942
Db	1438	AACATTTTACATGATTATAAAATTAATCAAAATTTTAAAGAGGTAAATCTCAAAAAACAAC	1379
Qy	1943	AAAAAAAAAAAAAA 1956	
Db	1378	ATATCAAAAGCAAA 1365	

RESULT 14

US-09-252-991A-12336/c
; Sequence 12336, Application US/09252991A

FACEBOOK NO. 8331793
GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

;
; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 12336

LENGTH: 489

TYPE: DNA

ORGANISM: *Pseudomonas aeruginosa*

US-09-252-991A-12336

Query Match 2.6%; Score 51.6; DB 4; Length 489;
Best Local Similarity 51.8%; Pred. No. 0.0015;
Matches 117: Conservative 0; Mismatches 109: Indels

1391	QY	GCTAAGCGCTGTTGAGGA	CTTGGAGCTGGAGAAATATTATTGAAC	TGCATTGATTGTGAT	1450
251	DB	GCACAAGCGCTCGAAGAGCGCTGGAGTGGGTGAGATTTTCTAAATTTCTATTGATCCGAGAT			192
1451	QY	GGTCAAGGAAAAAGGATTTCGATATAGATCTCGATCAAGCTAATATATCCGATGCTGTGAACATT			1510
191	DB	GGAGTGCAGAAAGGCTTCGACAACGCTCTAGTGGAAATATCGCTTCTAACGTCCTCATGTG			132
1511	QY	CTGTGTTATCCCAAGCAGCGGTGCAGGAGTCCGCTGATCAC	TCTCTCCGAAGTCTTTTAATGAA		1570
131	DB	CCAGTGNATGCCCTGTGGTGGAGCTGGCTCATCGCTGACCTCATCGATCTCTTTTGAGCGT			72
1571	QY	ACCAACGCATCTCTGTCGCCCTTTCGAGCTGGCATTTTCCATCGCAAG			1616
71	DB	ACGTGTGTGTGCGCAGTAGCGCGGGAAGGCTATTGTTTTTCAATG			26

RESULT 15

US-09-252-991A-12156

```

; Sequence 12156. Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: ASRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12156
; LENGTH: 798
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12156

```

```
Query Match      2.6%; Score 51.6; DB 4; Length 798;
Best Local Similarity 51.8%; Pred. No. 0.0019;
Matches 117: Conservative 0; Mismatches 109; Indels
```

1391	QY	GCTAAGCGTGTTCAGGAACCTTCGAGCTCGAGAAATATTATTGNACTGCATTGATTGTGAT	1450
514	Db	GCGCAAGCGCTCGAAGAGCTCGAGTGGGTGAGATTTTCTAAATTCTATTGATCGAGAT	573
1451	QY	GGTCAAGGAAAAGGATTTCGATATAGATCTGATCAAGCTAATATCCGATGCTGTGGAACATT	1510
574	Db	GGAGTCAGAAAGGCTTCGACAAAGCTCTAGTGGAAATATCGCTTCTAACGTCCTATGTC	633
1511	QY	CCTGTTTATCGCAAGCAGCGGTGCAGAGTGCCTGTGATCACTTCTCCGAAGTCTTTAATGAA	1570
634	Db	CCAGTGATCGCCTGTGTGGAGCTGSCCTCCATCGCTCACTCATCGATCTTTTGGCGGT	693
1571	QY	ACCAACGATCTGTGCCCTTCGAGCTGGGATTTTCCATCGCAAG	1616
694	Db	ACGTGTGTCTCGCAGTAGCGGCGGAAGCCTATTGTTTTTCATG	739

Search completed: January 20, 2004, 23:52:19
Job time : 141 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 23:50:02 ; Search time 698 Seconds
(without alignments)

9877.417 Million cell updates/sec

Title: US-09-831-233A-1

Perfect score: 1956

Sequence: 1 gcacgagaaggactggca.....aaaaaaaaaaaaaaaaaaaaa 1956

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2324096 seqs, 1762381658 residues

Total number of hits satisfying chosen parameters: 4648192

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	354	18.1	1506	12	US-10-369-493-46239
2	335.6	17.2	1656	12	US-10-369-493-25187
3	303.2	15.5	1732	12	US-10-369-493-27128
4	102.8	5.3	906	12	US-10-369-493-24907
5	98.4	5.0	819	12	US-10-369-493-45147
6	98.2	5.0	759	12	US-10-369-493-47005
7	95.8	4.9	1030	12	US-10-369-493-36928
8	92.2	4.7	816	12	US-10-369-493-24674
9	91.2	4.7	780	12	US-10-369-493-42098
10	91	4.7	768	12	US-10-369-493-43385
11	87.6	4.5	564	10	US-09-974-300-4849
12	84.4	4.3	405	10	US-09-974-300-4861
13	84.2	4.3	819	12	US-10-369-493-35008
14	84	4.3	759	12	US-10-369-493-41126
15	83.6	4.3	757	10	US-09-974-300-302

16	83.6	4.3	757	10	US-09-974-300-471	Sequence 471, App
17	82.6	4.2	636	10	US-09-974-300-364	Sequence 364, App
18	81.4	4.2	750	12	US-10-369-493-34025	Sequence 34025, A
19	81.4	4.2	753	12	US-10-369-493-40377	Sequence 40377, A
20	77.8	4.0	777	12	US-10-369-493-44423	Sequence 44423, A
21	75	3.8	6346	8	US-08-781-986A-237	Sequence 237, App
22	74.2	3.8	768	12	US-10-369-493-37473	Sequence 37473, A
23	72.4	3.7	768	12	US-10-369-493-31890	Sequence 31890, A
24	70.2	3.6	771	12	US-10-369-493-31035	Sequence 31035, A
25	68.6	3.5	774	12	US-10-369-493-24012	Sequence 24012, A
26	68.6	3.5	780	12	US-10-369-493-33621	Sequence 33621, A
27	68.4	3.5	771	12	US-10-369-493-28278	Sequence 28278, A
28	68.4	3.5	774	10	US-09-738-626-2291	Sequence 2291, Ap
29	68.4	3.5	1830121	15	US-10-329-360-1	Sequence 1, Appli
30	68.4	3.5	3309400	10	US-09-738-626-1	Sequence 1, Appli
31	67	3.4	792	12	US-10-369-493-34476	Sequence 34476, A
32	66.8	3.4	24417	15	US-10-216-209-1	Sequence 1, Appli
33	66.4	3.4	786	12	US-10-369-493-26426	Sequence 26426, A
34	65.4	3.3	640681	10	US-09-790-988-1	Sequence 1, Appli
35	65.2	3.3	750	12	US-10-369-493-34677	Sequence 34677, A
36	65	3.3	765	12	US-10-369-493-43797	Sequence 43797, A
37	65	3.3	783	12	US-10-369-493-40776	Sequence 40776, A
38	64.2	3.3	419	10	US-09-960-352-11234	Sequence 11234, A
39	63.4	3.2	765	12	US-10-369-493-39369	Sequence 39369, A
40	63.4	3.2	765	12	US-10-369-493-39760	Sequence 39760, A
41	63.4	3.2	768	12	US-10-369-493-39003	Sequence 39003, A
42	63.2	3.2	792	12	US-10-369-493-35791	Sequence 35791, A
43	62.8	3.2	768	12	US-10-369-493-31654	Sequence 31654, A
44	62.6	3.2	762	12	US-10-369-493-38811	Sequence 38811, A
45	62.6	3.2	780	12	US-10-369-493-35155	Sequence 35155, A

ALIGNMENTS

RESULT 1
US-10-369-493-46239
; Sequence 46239, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 46239
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-46239

Query Match	18.1%	Score	354	DB	12	Length	1506
Best Local Similarity	55.3%	Pred. No.	8.2e-77				
Matches	831	Conservative	0	Mismatches	640	Indels	33
Gaps	6						
QY	198	ACGCCCTTATCTTCTCGGCTTGGGCTTTGCACCTTCAATGGATGTTCTTATTCGTA	257				
Db	2	AGTGCTGATTTTCTCGTGTGGAACTTTGGTTTCGTATGCGATTCCTTAGCAAGC	61				
QY	258	AAAGGGCTGGCTGAAGCACTCTGTACTTACATTCAGAAATGATCGACCTTTCTGGGTATAT	317				
Db	62	AGGATTTTATAGACCTTGGCTGCTATGCTTTGTCTGGCAACCTTTTATGGCTGAT	121				
QY	318	GCGTGGGATTCAGCTACTCTTTGAGTCAAGTGAAGAAATGGTCCAATTCAGGCTTGG	377				

Db 122 GCGTTGGAATCCAGACATTAATTCGAGGGGCTCTGTTGAAGCACCACATCTTAAAGGGTTGG 181
Qy 378 GCTTGATTCCTGACGGGTTGGGGCTTTTGAATCATCCAAATGTTTAAAGGTCGCACATA 437
Db 182 GCGTATTTCCGGGGTTAGTCCAAAGGTTTGACAA---CGATGACAAGACCGTCCCTCATTA 238
Qy 438 TTGATGGCATGCTTGGATATAAAGAGGGTCAGCAATTTTAGATGATGTGGGAATC 497
Db 239 TTGGTTGGAATCTCTGTGCTGTACGATCTGACACTTCAAAAGAGTTTTTTTGGTATGGCC 298
Qy 498 AACATG-----TGATTTTGTTCACCTCATATCGAGCCAATGCGGAGGACAAACA----- 546
Db 299 CACATGACAAATTTTATTTTGTGCACTCTTATATGATTCAGAAAAGGGACTTATCCTTC 358
Qy 547 ----AGATGGATTTCACTACATGACGATATGG---TGACGATTTTATTTGCAATCCATTC 599
Db 359 CTCAGAGTTCAAAATTTGCTACACGAAATATGMAATGAAACCTTTGTAGGTGCTATTG 418
Qy 600 AGAAGGAATGTCATGCGAGTCCAAATTTTATCCCGAGAGAGTGGAGGTGTTGGACTTT 659
Db 419 TAAAGAAATATTTTGTAGCTACCCAAATTTTATCCGGAATAAGTGGATCTGCAAGTTTC 478
Qy 660 CCATATTGAGAAGATTTTGAATGCTGATTCCTTTAAACAAACAAAGACAGCAATGA 719
Db 479 GTTGTTTAAGGCATTTTAAACCGTAAATTTAGACCACTATAAGTGTGAAGCATCTA 538
Qy 720 ATGGAAGGCTCTAAACTTG-----CAAAGAGATTAATTTGCTTGGCTTGTATGTGA 770
Db 539 AACTTATAGAGAACTCTCTTGGCGGATTAACAAAGCGTATTAATTTGCTTGTCTAGATGTC 598
Qy 771 GGGCAATGATATGCGGATCTTGTGTAACCAAGGAGGACCAATATGATGTGAGAAC 830
Db 599 GTTCCACGACGCTGGTGACCTTGTAGTTACAAAGGTGATCAATATGATGTCCGTA 658
Qy 831 GTACAGAAGAGAAATGAGTCAAGAACTTGCAGAGCTTGTGAACCTTGTGGGCGAGTATT 890
Db 659 AATCTTCTGAGCGAGTTAGAAATCTTGGAAACCTTGTGGAGTTATGTCAAGATATT 718
Qy 891 ATTTAGACGGTGTGATGAGTCAAGTCTTCTTAAACATTTATGTTTCCGGGACTTCCCTC 950
Db 719 TTCAAGAAGGCGCTGACGAAAGTCTGTTTTTTTGAACATCACTTCAATTTAGAAATTTGTCAA 778
Qy 951 TAGCGGATCTACCATGCTCAGAGTCTTGCAACGCGATCTGAAAAGCTTTTGTGCGCAT 1010
Db 779 TGGCGGATGCACTATGTTACAGGTTTTTGAAGAGGCGAGCTCAAAACAGTATTTGTACCTT 838
Qy 1011 TAACTGTCCGGGTCGCATCAGGATTTTACTGATGCAAAATGGAAGGTATTTTCTAGTC 1070
Db 839 TAACAGTAGGTGGTGGATCAGAGATGTATCAGATCTCTGATGGGACTTTTCACTCCTG 898
Qy 1071 TAGAAGTGGCTTCAGATATTTTCAAGATCGGGCGCGATAGAGTTTCGATCGGAAGTGATG 1130
Db 899 TCGAAGTTGCTGGTATTTACTTTAGATCAGGTGCGGACAAAGTATCTATTGGAAGTGACG 958
Qy 1131 CAGTTTACACTGCTGAGGAATATATTAACCCGAGTGAA---CACAGGAAGAGCAGCA 1187
Db 959 CCGTTTATGCGCGAGAAAGATTTACGAAAATGGAAGAAATGAGTGGAAACATCTGCTA 1018
Qy 1188 TAGAGCAGATATCTACAGTATATGCTAAACAGGCGAGTGGTTGTAAAGCAATGATCCTCGCC 1247
Db 1019 TAGAAACAATTTCTAAGGCCCTTAGTATCAGGCAGTTGTTATTTCCGTTGATCTTAAC 1078
Qy 1248 GAGTTTACTTGAGAAAACCGGATGAAGTATTAAGCAATTAAGCAATCAAAAGTAAAGCCATCCAG 1307
Db 1079 GTCAGTATGTAAGATTTCTTGAAGATATAACCAACCATGTTGTGAAAACATCTCGGCTTG 1138
Qy 1308 GTCCAAACGGTGAAGATATGCTGATCAGTCACTGTTTAAAGTGGAGGAGGGA 1367
Db 1139 GACCTAATGGAAGACATATTTGTTGGTATCAGTCAACCGTTAAAGGAGGTGCGTAATC 1198
Qy 1368 GACCCATCGGAGCTTATGAACCTAGCTAAGGCTGTTGAGGAACTTTGAGGCTGGAGAAATAT 1427
Db 1199 GTGATATTTGTTGTTGAACTTACTCGTGTCTGTGAAGCAATGGGTGCTGGTGAAGTAC 1258

Qy 1428 TATTGAACCTGATTTGATGGTCAAGGAAAAGGATTCGATATAGATCTGATCAAGC 1487
Db 1259 TATTAAATTTGATGGATCAAGACGGTTCAATGCTGGGTATGACATAGAACTTGTTCGTC 1318
Qy 1488 TAATATCCGATGCTGTGAACATTTCTGTGTTATCGAACAGCGGTGAGGAGTGCCTGATC 1547
Db 1319 TTGTTAAAACCTCTGTGAATATTTCCAGTAAATAGCTTCAAGTGGTGTGTTATCCCAAC 1378
Qy 1548 ACTTCTCCGAAGTCTTTTAAATGAAAACCAACGATCTCTGCTGCCCTTGAGCTGGCATTTTC 1607
Db 1379 ATTTTGAAGAAGTATTTTAAAGAAAACGATTTGTGATGACGCTCTTGCCGCGGAATTTTC 1438
Qy 1608 ATCGCAAGAGGTTCCTCAATTTAAGGCTGTTAAAGAGCACTTTGTTGAAGGAGGCAATTGAAG 1667
Db 1439 ATCGCAAACTTGTGCGATAGAACGCTCAAGAAATATCTGGCAATACACGAGTTCTGG 1498
Qy 1668 TTAG 1671
Db 1499 TTCCG 1502

RESULT 2

US-10-369-493-25187
; Sequence 25187, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 25187
; LENGTH: 1656
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-25187

Query Match 17.2%; Score 335.6; DB 12; Length 1656;
Best Local Similarity 60.1%; Pred. No. 3.1e-72;
Matches 576; Conservative 0; Mismatches 379; Indels 3; Gaps 1;

Qy 730 TTCTAAACTTGGAAAGAGATTAATTCCTTGCCTTGATGTCAGGGCAAAATGATTAATCGGGA 789
Db 699 TTATGCTCTAACCGCAGAAATTAATTCCTTGTCTGTGATGACGTACTAATGACCAAGTGA 758
Qy 790 TCTTGTGTTAAACCAAGGAGACCAATATGATGTGAGAGAACGTACAGAGAGAGATGAGGT 849
Db 759 TTTGGTGTGTACTTAAGGTGATCAATCGATGTACGTGAAAAGTGTGTTAAAGTGT 818
Qy 850 CAGAAACCTTGGCAAGCCTGTTGAACCTTGTGGCCAGTATTTATTTAGACGGTCTGTATGA 909
Db 819 TAGAAACCTTGGTAAAGCTGTTCAAGTTGGCACAGAAATATTAACCAACAGGTCGGATGA 878
Qy 910 GGTCACTTCTTAAACATTTACTGTTTCCGGGACTTCCCTCTAGGCGATCTACCCATGCT 969
Db 879 AGTAAACATTTTGAATATAAATCTCTTTTAGAGATTGTCTTTTGAAGGATACTCCGATGCT 938
Qy 970 ACAGGTCTTGCACACGCGCATCTGAAAAACGTTTTTGTGCCATTAATCTGTCCGGGGTGGCAT 1029
Db 939 AGAGGTCTGAAACAGCCCAAGACAGTCTTTTGTTCATTCACAGTCCGTGGGGGAT 998
Qy 1030 CAGGGATTTTACTGATGCAAAATGGAAGGTATTTATTTAGTCTAGAGAGTGGCTTCAGAGTA 1089
Db 999 CAAGGATTTGTCGATGTTGATGAAACCAAAATACCTGCTTTTGAAGATTGCAAGTCTATA 1058

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 24907

; LENGTH: 906

; TYPE: DNA

; ORGANISM: Methanobacterium thermoautotrophicum

US-10-369-493-24907

Query Match 5.3%; Score 102.8; DB 12; Length 906;

Best Local Similarity 55.2%; Pred. No. 7.7e-15;

Matches 200; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 1316 GGTGAGGAATATGCTGGTATCAGTGCATCTGTTAAATGGTGGAGGAGGAGACCCATC 1375

DB 544 GATGATGGATCTGCTGGTACGAGTGCAGCATCTATGGTGGAGAGATTTCACAGGTATT 603

QY 1376 GGAGCTTATGAATAGCTAAGCTGTTTGAGGAACCTTGGAGCTGGAGAAATATTATTGAAC 1435

DB 604 GACGCGTGAATGGGCAATGGAAATGCCAGATAGAGTGCAGGTGAAATTTCTCTCACA 663

QY 1436 TGCATTGATTGTGATGCTCAAGGAAGGATTGATATAGATCTGATCAAGCTAATATCC 1495

DB 664 TCAATGACCGCTGACGCTACAAAGATGGGCTACGACATTCCTCCACAGGACCATGAGC 723

QY 1496 GATGCTGTGAACATTCCTGTTATCGCAAGCAGCGGTGACAGTCTGCTGATCACTTCTCC 1555

DB 724 GAGAACCTTGACATACCTGTGATCGCATCAGGGGTGAGGAACTGGAACATCTAT 783

QY 1556 GAAGTCTTTAATGAAACCAACGCATCTGTCCTTGCAGCTGGCACTTTTCCATCGCAAA 1615

DB 784 GAGGCAATTCACAGATGTTAAGGCAGACGCTGCATCTGCTCAAGCATATTCCTCAAT 843

QY 1616 GAGTTCCAAATTAAGCTGTTAAAGACCATCTGTTGAGGAAGGATTGAAGTTAGATTG 1675

DB 844 GAATACCTTCTCCCTGAGTGAAGGATGATCTCAGATCAAGGGGAGTTCCCATAGATTG 903

QY 1676 TA 1677

DB 904 TA 905

RESULT 5

US-10-369-493-45147

; Sequence 45147, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 45147

; LENGTH: 819

; TYPE: DNA

; ORGANISM: Methanococcus jannaschii

US-10-369-493-45147

Query Match

Best Local Similarity 5.0%; Score 98.4; DB 12; Length 819;

Matches 219; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 1259 AGAAACCCGATGAAGTAGAATTTAAAGCCATCAAAAGTAAGCCATCCAGTCCAAACGGT 1318

DB 397 AGACACTATGTTAATGAAGATGAGATAGATAAATAAATAATGCTGTTAAAGTAGAG 456

QY 1319 GAGGAATATGCTGGTATCAGTGCATCTGTTAATGGTGGAGAGAAAGGAGACCCATCGGA 1378

DB 457 GATGGTTATTGCTGGTTTGAAGTTTATATATACGAGGAGGAGAAAGAAACAGGTATAGAT 516

QY 1379 GCTTATGAACCTAGCTAAGCTGTTGAGGAACCTTGGAGCTGGAGAAATATATTGAAGTGC 1438

DB 517 GCCATAAAGCTGGGCTAAAAAGTTGAAGAAATTTGGGAGCTGGAGAGATTTTATTGCAAGT 576

QY 1439 ATTGATTGTGATGCTCAAGGAAAAAGGATTCGATATAGATCTGTATCAAGCTAATATCCGAT 1498

DB 577 ATTGATTAAGATGGGCAAAAAAGTGGCTATGATTTGATATTGCAAAAGGAGATTTCTAAA 636

QY 1499 GCTGTGAACATTCCTGTTATCGCAAGCAGCGGTGACAGGATCGCTGATCACTTCTCCGAA 1558

DB 637 AGTGTAAACTCCCTGTTATTGCAAGTGGAGGTTGTGAAACCCAGAACATGTTTATGAG 696

QY 1559 GTCTTTAATGAACCAACGCATCTGCTGCCCTTGCAGCTGGCATTTTCCATCCCAAGAG 1618

DB 697 GCATTTTGTATGGAAGGACGATGCCGCATTTAATGCGCAGGGGATATTGCACTATAGAGAA 756

QY 1619 GTTCCAATTAAGCTGTTAAAGAGACACTTGTGTAAGGAAAGGGATTCGAAGTTAGATTGTA 1678

DB 757 TATACANTAGAAGATAAAAAATACTGTGCTGATAGAGGAATACCTATGAGATTATTA 816

RESULT 6

US-10-369-493-47005

; Sequence 47005, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Cao, Yongwei

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 47005

; LENGTH: 759

; TYPE: DNA

; ORGANISM: Bacillus subtilis

US-10-369-493-47005

Query Match 5.0%; Score 98.2; DB 12; Length 759;

Best Local Similarity 52.8%; Pred. No. 9.5e-14;

Matches 234; Conservative 0; Mismatches 208; Indels 1; Gaps 1;

QY 1232 AGCATTTGATCTCGCCGAGTTTACTTGAGAA-AACCCGATGAAGTAGAATTTAAAGCCAT 1290

DB 315 AGCTGTTCTCCGCCCTGAGCTGATTACAGAAGGAGCGGAGTTTTTCGTTTCTCAATGCAT 374

QY 1291 CAAAGTAAGCCATCCAGTCCAAACCGTGGAGGAATATGCTGCTATCAGTGCATGTTAA 1350

DB 375 CGTACTTGGCAATGATGCCAAGTATGCAAAAGAACTTGACACATATAAGGCTTACAGCA 434

QY 1351 TGGTGGACGAGAAGGAGACCCATCGAGCTTATGAAGTAGTAAAGCTGTTGAGGAAC 1410

DB 435 CGCGGCGAGAAAGAAACAGATTGGGAGGTCACCGCGTGGCAAAAGAGCGCTCAACG 494

QY 1411 TGGAGCTGGAGAAATATTATTGAACTGCAATGATTGATGGTCAAGGAAAAAGGATTGCA 1470

DB 495 CGGGGCGAGGAGAAATATTGCTGCAAGCATGGACTCCGACGGTGCAGAAAAAAGGTTTCGA 554


```
; SEQ ID NO 42098
; LENGTH: 780
; TYPE: DNA
; ORGANISM: Lactococcus lactis
; US-10-369-493-42098

Query Match      4.7%; Score 91.2; DB 12; Length 780;
Best Local Similarity 53.0%; Pred. No. 5.2e-12;
Matches 195; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 1307 GGTCCAAACGGTGAGGAATATGCTGTATCAGTGCACATGTTAATGTTGACGAGAGGG 1366
Db 391 GCTAAAAACGAGCAGACCATAGGGATATGATGCTATATCAAAAGTGGCGGTGAAAT 450
QY 1367 AGACCCATCGGAGCTTATGAACCTAGCTAAGGCTGTTGAGGAACCTTGAGCTGAGAAATA 1426
Db 451 GCAGGCTTCACTTACTAGTTGATGGCTTAAAGTGTGAAGATTTAGGAGCAGCGCAAT 510
QY 1427 TTATTGAACCTGATGATGCTGATGTCGCAAGGAAAGGATTCGATATAGATCTGATCAAG 1486
Db 511 TTACTAACCTCAATGATAGGATGGAACAAAGCAGGCTACGATTTGGAATGCTGAT 570
QY 1487 CTAAATFCCGATGCTGTGAACATTCCTGTATTCGCAAGCAGCGTGCGAGATGCTGAT 1546
Db 571 GAGCTTTGTACTGCTGTATATTCCTGTGTAGCAAGTGGCGTGTGGGAAGATTCG 630
QY 1547 CACTTCTCCGAAGCTTTAATGAACCAACGCATCTGCTGCCCTTGCGAGCTGCAATTTTC 1606
Db 631 GACATTTGTAAGTTTTTCAAAATACAAAGAGCGATGCTGCCCTTGTGCTTCAATTT 690
QY 1607 CATTGCAAGAGGTTCCAAATTAAGGCTGTTAAAGAGCACTTGTGGAAGGAGGATTTGAA 1666
Db 691 CATTATGGCGAAGCAACAGTGTGATGAAGTTAAAGACGAACCTTATAAAAAATATATCCA 750
QY 1667 GTTAGATT 1674
Db 751 GCAAGAT 758

RESULT 10
US-10-369-493-43385
; Sequence 43385, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 43385
; LENGTH: 768
; TYPE: DNA
; ORGANISM: Nitrosomonas europaea
; US-10-369-493-43385

Query Match      4.7%; Score 91; DB 12; Length 768;
Best Local Similarity 51.9%; Pred. No. 5.8e-12;
Matches 205; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 1281 TTAAGCCATCAAGTAAGCATCCAGTCCAAACGGTGAGGATATGCTGTATCAGT 1340
Db 374 TGATCGCAATCGATGCCAGACAAATCTGATGCAAAATCTCGAATCTCCAGCTGGGAAG 433
QY 1341 GCACCTGTAATGTTGGAGCAGAGAGAGAGCCATCGGAGCTTATGAACCTAGCTAAGGCTG 1400

; SEQ ID NO 42099
; LENGTH: 780
; TYPE: DNA
; ORGANISM: Lactococcus lactis
; US-10-369-493-42099

Query Match      4.5%; Score 87.6; DB 10; Length 564;
Best Local Similarity 51.6%; Pred. No. 3.3e-11;
Matches 263; Conservative 0; Mismatches 229; Indels 18; Gaps 2;

QY 179 GAGGATATTTAAATGCTAAGCGCCTTATCTTCTCGGCTGGGCGCTTTGACCTGCA 238
Db 45 GAAGAATTGGCAAGGCCAAGCGCTTGATCCCTCGGGGTGCGCGCTTTCTCTGACGCA 104
QY 239 ATGGATGTTCTTATTCTGTAAGGGCTGGCTGGAAGCACTCTGTACTTACATTCAGAATGAT 298
Db 105 ATGGATATTTCTACAAACAAACGTCAGCCCTTTTACACAACTGGGTAGCGGAAAC 164
QY 299 CGACCTTTCTGGGTATATGCTCGGATTCAGCTACTCTTTGAGTCAAGTGAAGAAAAT 358
Db 165 AAACCATTTACTCGGCATTTGCCCTTGGGATGCAACTATTGTTTGGTCAAGCAGGAAACAT 224
QY 359 GGTCCCAATTCAGGCTTTGGCTTGATTCCTGGAGGGTGGGGGTTTGAATCATCCA-- 416
Db 225 CAAACCAACAGCAGCGCCTTGGTTACTTGGCTGGAGCGGTGCAACGTTTTCTGGAACGACG 284
QY 417 -----ATGCTTTAAGGGTGCCACATATTGGATGGCATGCTCTGGATATAAAGGAA 466
Db 285 GCAGAGGGGACCGCTACAAAGTCCGCATATGGGATGGATGAATAGCTTGAATCCATAGG 344
QY 467 GGGTCAAGCAATTTTATGATGATGTTGGGAATCAACATGTGTATTTGTTTCTACTCATATCGA 526
```

Db 345 CCCACACCACTAACAGAGGTGTTCTCGAAGGCTACGTATATTTTGTCCATTCTATGTC 404
Qy 527 GCCAATGCCGAGGACAAAGAGTGTGATTTCTATCATCGACGCTATGTCGAGATTTT 586
Db 405 GTCGTTACAG-----AAACGGACATCATCGTTCCAGACGAGTATTTACCAAGTGT 458
Qy 587 ATTGCATCCATTTCAGAAAGGAATGTTTCATGCGAGTCCAAATTCATCCCGAGAGAGTGG 646
Db 459 CTTGCTATTGTGCAAAAAGGCAAGTGTATGGCATGCAATTCATCCCGAAGAAAGCAGC 518
Qy 647 GGTGTTGGACTTTCATATTGAGAGATTT 676
Db 519 ACTGTGGCATGCGACTGCTGAAACGCTTT 548

RESULT 12

US-09-974-300-4861
; Sequence 4861, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4861
; LENGTH: 405
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-4861

Query Match 4.3%; Score 84.4; DB 10; Length 405;
Best Local Similarity 54.5%; Pred. No. 1.7e-10;
Matches 169; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
Qy 1350 ATGGTGACGAGAGGAGAGCCCATCGAGCTTATGAACCTAGCTAAGGCTGTTGAGGAAC 1409
Db 82 ATGGTGCCCAACACAGACTGACTGTGTTAGGACCGATTGGGCCAAGGAGGCTGTCCGCT 141
Qy 1410 TTGGAGCTGGAGAAATATTATGAACCTGCAATTTGATGATGTCATGAAGGAAAGGATTCG 1469
Db 142 TAGGGGCGCGGAACCTGCTGTAACCTAGCATGATCAAGATGAGCGAGGAGGAGGTTG 201
Qy 1470 ATATAGATCTGATCAGCTAATATCCGATGCTGTGAACATCTCTGTTATCGAAGCAGCG 1529
Db 202 ATTTGTGTTGTTAAAGCGCATCAATGAAGTAGTACAGAGTGGCCCGGATTCCTCTGGTG 261
Qy 1530 GTGCAGGAGTGGTGATCACTTCTCCGAAGCTCTTTAATGAAACCAACGCACTGCTGCC 1589
Db 262 GAGCAGGAGCGGCAAGGCACTTTGTTGATGTTTCAACACATATTACGAGATGACGCC 321
Qy 1590 TTGCAGCTGGCATTTTCCATCGCAAGAGGTTTCCAATTAAGGCTGTTTAAAGACACTTGT 1649
Db 322 TTGGGCGCATGATTTTCCATTACAAAGAACGTCATTTCAAGAAGTAAAGCCCGTTTAC 381
Qy 1650 TGAAGNAGG 1659
Db 382 GGAAGAAGG 391

RESULT 13

US-10-369-493-35008
; Sequence 35008, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 35008
; LENGTH: 819
; TYPE: DNA
; ORGANISM: Methanosarcina mazei
US-10-369-493-35008

Query Match 4.3%; Score 84.2; DB 12; Length 819;
Best Local Similarity 53.9%; Pred. No. 2.9e-10;
Matches 173; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
Qy 1327 TGCCTGTTATCATGTCACCTGTTAATGTTGACGAGGAGGAGCCCATCGAGCTTATGA 1386
Db 471 TGCCTGTTATGAGGTTGTAAATTTACGAGGCGACAGAGGCTACAGGAATCGATCCCGTCA 530
Qy 1387 ACTAGCTAAGGCTGTTGAGGAACTTGGAGCTGGAGGAAATATTATGNACTGCAATTGTTG 1446
Db 531 GTGGCAAGAAAGCAGAGAATTTGGGTTCCGGAGAAATCTCTCTCACAAGCATGGACCG 590
Qy 1447 TGATGTCGAAGGAAAGGATTCGATATAGATCTGTGATCAAGCTAATATCCGATCTGTGAA 1506
Db 591 GGATGTCACGCTGTGCGGTTATGACCTGCCAATTACGAAAAAGCTCTCTGAAGAGCTCGA 650
Qy 1507 CATTCCTGTTATCGCAAGCAGCGGTCGAGAGTGGCTGATCACTTCTCGAAAGTCTTTAA 1566
Db 651 TATCCCGATAATAGCTTTCAGGAGGTCGCGAAACCTCTCAGCACATTTTACGAAGGATTTTC 710
Qy 1567 TGAACCAACGATCTGTCCTTCAGCTGCGATTTTCCATCGCAAGAGGTTTCCAAT 1626
Db 711 AATTGGAAGGCGATGCTGCACTTCGACGAAGTATCTTCCACTTCAGAGAGTACTCAAT 770
Qy 1627 TAAGGCTGTTAAAGACACTT 1647
Db 771 TAAGGAAGTCAAGAATACTT 791

RESULT 14

US-10-369-493-41126
; Sequence 41126, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 41126
; LENGTH: 759
; TYPE: DNA
; ORGANISM: Bacillus halodurans
US-10-369-493-41126

Query Match 4.3%; Score 84; DB 12; Length 759;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 21:40:56 ; Search time 4247 Seconds
(without alignments)
11193.679 Million cell updates/sec

Title: US-09-831-233A-1
Perfect score: 1956
Sequence: 1 gcacgagaggaactggcga.....aaaaaaaaaaaaaaaaaaaaa 1956

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estnu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_hc:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_hc:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: em_gss_hum:*
 - 18: em_gss_inv:*
 - 19: em_gss_pln:*
 - 20: em_gss_vrt:*
 - 21: em_gss_fun:*
 - 22: em_gss_mam:*
 - 23: em_gss_mus:*
 - 24: em_gss_pro:*
 - 25: em_gss_rod:*
 - 26: em_gss_plg:*
 - 27: em_gss_vtl:*
 - 28: gb_gss1:*
 - 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	750.4	38.4	2142	11	AY109594 Zea mays
2	498	25.5	780	10	BF274106 GA_Eb002
3	428.4	21.9	638	13	BU046804 PP_LEa002
4	425.6	21.8	668	12	BM878913 P9-G02 Sw

5	335.8	17.2	727	13	BQ112714
6	332.4	17.0	600	12	BG833389
7	330.8	16.9	739	9	AL820092
8	327.2	16.7	533	10	BE340229
9	317.2	16.2	560	10	BF519112
10	311	15.9	729	14	CD453914
11	308.2	15.8	591	12	BM086170
12	301.4	15.4	716	13	BQ869893
13	300.8	15.4	1041	29	CNS06GQF
14	298.4	15.3	569	10	BE353528
15	289	14.8	511	12	BE833390
16	286.8	14.7	434	14	CA925771
17	276	14.1	645	10	BF053343
18	274.2	14.0	562	9	AI995084
19	272.4	13.9	639	13	BQ116797
20	262	13.4	615	9	AW694763
21	254.4	13.0	626	13	BQ578659
22	253.4	13.0	625	13	BQ578541
23	252	12.9	661	9	AU239364
24	251.2	12.8	562	10	BE920620
25	249.2	12.7	460	10	BE999461
26	248.4	12.7	602	13	BQ992987
27	247	12.6	602	13	BU005663
28	245	12.5	570	10	BE449432
29	245	12.5	642	9	AI055415
30	243.8	12.5	595	13	BQ982164
31	240	12.3	589	10	BF050974
32	236.6	12.1	639	9	AW066760
33	235	12.0	386	10	BE523897
34	232	11.9	501	14	H37732
35	230.6	11.8	494	12	BG833391
36	226	11.6	645	10	BE341560
37	224	11.5	598	13	BU971680
38	217.6	11.1	540	9	AJ470077
39	215.8	11.0	557	9	AV546666
40	205.6	10.5	334	10	BG239122
41	201.8	10.3	898	29	CNS06KXT
42	194.4	9.9	544	14	CD448587
43	192.2	9.8	583	13	BQ588079
44	190	9.7	483	10	BF704745
45	189.2	9.7	476	12	BM065703

ALIGNMENTS

RESULT 1	AY109594	AY109594	2142 bp	mrna	linear	HTC 17-OCT-2002
LOCUS	Zea mays	CL287_1	mrna sequence.			
DEFINITION	Zea mays	CL287_1	mrna sequence.			
ACCESSION	AY109594					
VERSION	AY109594.1	GI:21213374				
KEYWORDS	HTC.					
SOURCE	Zea mays					
ORGANISM	Zea mays					
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
AUTHORS	Hayney,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitlitt,M.S.,					
	Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.					
TITLE	Maize Mapping Project/DuPont Consensus Sequences for Design of					
	Overgo Probes					
JOURNAL	Unpublished (2002)					
REFERENCE	2 (bases 1 to 2142)					
AUTHORS	Coe,E.H.					
TITLE	Direct Submission					
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of					
	Missouri, Columbia, MO 65211, USA					
COMMENT	If you are interested in getting corresponding physical clones,					
	these are publicly available from ZmDB and may be found by BLAST					
	searching at MSL, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR,					
	www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the					

maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB:

FEATURES

Location/Qualifiers
 1. 2142
 /organism="Zea mays"
 /mol_type="mRNA"
 /db_xref="maizeDB:611206"
 /db_xref="taxon:4577"
 /clone_lib="Maize Mapping Project/DuPont Cornsensus Library"
 /note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"

BASE COUNT 477 a 467 c 505 g 479 t 214 others
 ORIGIN

Query Match 38.48; Score 750.4; DB 11; Length 2142;
 Best Local Similarity 64.9; Pred. No. 1.5e-82;
 Matches 1068; Conservative 0; Mismatches 569; Indels 9; Gaps 2;

52 CGCCGCTTGGCTGCGAGGATCTCTGGTGAATTTACTTGAATACGGTCTGG 111
 Db CGTCTCGTCCGTGCTCTCGGCGCAACACGGTGACTCTGCTGACTACGGCGGG 349
 112 AAATGTTAGAGTGTGAGAACCGCATCCGCACATTTGGATTTGATATCAAGATGCA 171
 Db GAACGTACGACGGTGGCAAGCAATTCGCTACCTCGGCTTGCATCCGACGTCGA 409
 172 AAGCCAGAGATATCTAAATGCTAAGCGCTTATCTTCTCGGCGTGGGCGCTTGC 231
 Db GAGCCCGAGGACATCTCGCGCGAACCGCTCTCTTCCCGGTGTCGGCGCTTCGG 469
 232 ACCTGCAATGATGCTCTTATTCGTAAGGCTGGCTGAAGCACTCTGTACTTACATTC 291
 Db CTCGCCATGAGCGTCTCTACAGGACGGGATGGCCACGCACTCGTGATACATCCA 529
 292 GAATGATCGACCTTTCTCGGTATATCCCTGGGATTCAGCTACTCTTTGAGTCAAGTGA 351
 Db AAGGACCGCCCTTCTAGGCACTCTCGCTCGGCTCCAGCTCTCTTCGATCCAGCGA 589
 352 AGAAATGGTCAATTCAGGCTCTGCTGATTCCTGGAGGCTGGGCGTTTGTATC 411
 Db GGAGAACGGCCCGGTGAGCGGCTCTCGGTGATACAGGCGTGGTCAAGCGATTCGACTC 649
 412 ATCCAAATGGTTAAGGTTGACACATATTTGGATGGCATGCCCTTGGATATAAGGAGGTC 471
 Db CTCAATGGCTCTAGTTCCACATATTTGGCTGGAACGCTCTCCAGATCACAGGACAC 709
 472 AGCAATTTTATGATGATGGGGAATCAACATGTTATTTTGTTCACCTCATATCG---AGC 528
 Db ACCACTGTTCAGGAGCTGATGCGCAGCATGTGCTACTTTGTTCACTCTACCGCTACT 769
 529 CAATGCGGAGGACAAAGAGTGGATTCATCTACATGACGATGCTGTCAGCATTTTAT 588
 Db GCCTTCAGATGCTAATAGAACTGGATTTCTCCATATGCAACTATGCTGACAGCTTTAT 829
 589 TGCAATCCATTAGAGGGAATGTTGATGATGCTCAATTTTCATCCCGAGAGGAGTGAGG 648
 Db ATCTCCATCTCAATGGGCAACATTCAGGCGATTCATTTTCCAGAAAGAGTGGAGC 889
 649 TGTGAGCTTTCCATATTTGAGAGATTTTGAATGCTGATTTCTTTTAAACACAAAGACA 708
 Db TACCGGACTTCTATTTAAAGAAATTTCTCAGACCTAACTCTCTTGGGACCA-----A 943
 709 GAAGCCATGATGAGAAAGCTTCTAACTTGCAGAGAGATTAATGCTTGCCTTGTATGT 768
 Db GGTCCCGACGAGTGAAGAGCATCAAAACTTTGCGAAGAGAGTATAGCATGCTTGTATGT 1003
 769 GAGGGCAATGATAATGGGGATCTTTGTTGTAACCAAGGAGGACCAATATGATGTGAGAGA 828

1004 TCCTTCAAAATGATAAAGGAGATCTTGTGTAAACAAAGCGCATCAATATGATTAAGAGA 1063
 Qy ACCTACAGAAGAGATGAGGTGAGAAACCTTGGCAGGCTGTTGAATCTTCTGGGAGTA 888
 Db TCATATAGCAGCAAGAGGTAAGAAACCTTGGCAGGCTGATTTAGCAAGCAGTA 1123
 Qy TTATTTAGACGGTGTGATGAGGTGAGCTTCTTAAACATTTACTGTTTCCGGGACTTCCC 948
 Db CTACATAGACGGTGTGATGAGGTGAGCTTCTTGAATATAAATGTTTCCGTCGACTTCC 1183
 Qy TCTAGCGATCTACCATCTACAGCTCTTGAACCGCATCTGAAACGTTTCTGTTGTC 1008
 Db ATTGGGTGATTTGCCAATGCTAGAGGTACTGCGTCTGCTCTGAAAGGNNNNNGTCC 1243
 Qy ATTAATCTGTCGGGGTGGCATCAGGATTTTACTGATCAAAATGGAAGGTATTTCTAG 1068
 Db ACTTACAGTTGGTNNNNNATACGAGACTTTCAGATGCAAAATGGAAGTACTACTCAAG 1303
 Qy TCTAGAAGTGGCTTTCAGAGTATTTTCAGATCGGCGCGCATTAAGTTTTCGATCGAAGTGA 1128
 Db TTTGGAGGTAGCATCAGATATTTTCAGTCCGCTGCTGACAAATTTCAATTTGGAAGTGA 1363
 Qy TGCAGTTTACCTGCTGAGGAAATATATAAACCAGGAGTGAAGACAGGAAAGACAGCAT 1188
 Db TGCCTGTTTATGCTGTAAGCCNNNNNACAGACTGCTGTAAGACAGGGAAGACAGCTT 1423
 Qy AGAGCAGATATCTACAGTATATGTTAACGAGGCTGTTGTAAGCATTTGATCTCGCG 1248
 Db GGAGCAAACTCTAGAGTATATGCGAATCAGGCTGAGTGTGTCAGTATTTGATCCTCGAGC 1483
 Qy AGTTTACTTTCAGAAAACCGATGAACTAGAAATTTAAAGCCATCAAAATGTAAGCATCCAGG 1308
 Db GGTATATGTCAAAGTCAAGAGATGTCATTTAAACTGTAAAGGTGTCCTACTAAAGG 1543
 Qy TCCAAACGGTGAAGAAATATGCTGATCAGTCACTGTTAATGTTGGTGAAGAGGAG 1368
 Db TCCATCGGAGAGAAATANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 1603
 Qy ACCCATCGGAGCTTATGAAGCTGTTGAGGACTTGGAGCTTGGAGAAATATTT 1428
 Db NNN 1663
 Qy ATTGAATCGATGATGTTGATGCTCAAGGAAAGGATTCGATATAGATCTGATCAAGCT 1488
 Db TCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1723
 Qy AATATCCGATGCTGTAACATCTCTGTTATTCGCAAGCAGCGGTGCAAGGAGTCTGATCA 1548
 Db GGTTCCTGATGCTGACAAATCCCTGCTTTCGAGCAGTGGTGGAGCTGTTCAACA 1783
 Qy CTCTCTCGAAGCTTTTAATGAACCAACGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1608
 Db NNNNNCTGAANN 1843
 Qy TCCCAAGAGGTTCCCAATTAAGGCTGTTAAAGAGCACTTGTTCGAAGGAGGATTTGAAGT 1668
 Db CCGAAGAGGTTTCTTATCTAGTGAAGAGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1903
 Qy TAGATTGTAAGGCGAGAAATCCTTGG 1694
 Db CAGGGTGAACGAGGAGATCCTTCGG 1929

RESULT 2

BF274106
 LOCUS GA_Eb020C04f Gossypium arboreum 7-10 dpa fiber library Gossypium
 DEFINITION
 accession cDNA clone GA_Eb020C04f, mRNA sequence.
 ACCESSION
 BF274106
 VERSION
 BF274106.1
 KEYWORDS
 EST.
 SOURCE
 Gossypium arboreum

ORGANISM
Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.

REFERENCE
1 (bases 1 to 780)

AUTHORS
Wang,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry D., Wood,T.C., Leslie,A. and Walkins,T.A.

TITLE
An integrated analysis of the genetics, development, and evolution of the cotton fiber

JOURNAL
Unpublished

COMMENT
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
High quality sequence stop: 726.

FEATURES
source
1..780
/organism="Gossypium arboreum"
/mol_type="mRNA"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Eb0020C04f"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/clone_lib="Gossypium arboreum 7-10 dpa fiber library"
/notes="Vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI"

BASE COUNT
231 a 123 c 206 g 220 t

ORIGIN
Query Match 25.5%; Score 498; DB 10; Length 780;
Best Local Similarity 79.2%; Pred. No. 1.3e-51;
Matches 616; Conservative 0; Mismatches 160; Indels 2; Gaps 2;
QY 881 GGCGAGTATTATTAGACGGTGTGATGAGTGTGAGTCTTAAACATTACTGTTCCGG 940
DB 1 GGACAGTATTACAAAGTGGGGCGATGAGTGTGAGTCTTAAACATTACTGTTCCGT 60
QY 941 GAATTCCTCTAGGCGATCTACCATGCTGAGTCTTTCGACGCGCATCTGAAACGTT 1000
DB 61 GATTTCCATTAGCGGATTTACCAATGTTGCGAGTATTAGGCACACATCAGAAATGTT 120
QY 1001 TTGTGCGCATTAATCTGCGGGGTGGCATCAGGATTTTACTGATGCAAAATGGAAGTAT 1060
DB 121 TTGTGCGCGTAAACAGTTGGAGGTGGTATACGGGATTTTACTGATGCAAAATGCGAGTAC 180
QY 1061 TATTTAGTCTAGAGTGGCTTCAGAGTATTTTCAGATCGGGCCCGATAGGTTTCGATC 1120
DB 181 TATTTAGTCTAGAGTGGCTTCAGAGTATTTTCAGATCGGGCCCGATAGGTTTCGATC 240
QY 1121 GGAAGTATGAGTCTTACCTGCTGAGGAATATTAAACCGGAGTGAAGACAGAAAG 1180
DB 241 GGAAGTATGAGTCTTACCTGCTGAGGAATATTAAACCGGAGTGAAGACAGAAAG 300
QY 1181 AGCAGCATAGACGAGATATTACAGTATATGGTAACCGAGCGTGGTTGTAAGCATTTGAT 1240
DB 301 AGCAGCTTAGACAAATTTCTAAAGTTTATGGAATTCAGGAGTGGTTGTAAGCATTTGAT 360
QY 1241 CCTCGCGAGTTTACTTGAGAAACCGATGAGTATTTAAAGCCATCAAGTAAGC 1300
DB 361 CCTCGTAGGTGATGTTTCAAGTCTTAATGATGTGAGTTCAGGATACACGAGTACCGGTC 420
QY 1301 CATCCAGTCCAAACCGTGAAGGAATATGCTGGTATCAGTGCATGTTAATGTGTGACGA 1360
DB 421 AAACAGGTCCAAATGGAAGGAATATGCTGGTATCAGTGTACGGTTAATGTGTGACGA 480
QY 1361 GAAGGGAGACCCATCGAGCTTATGAACACTAGCTAAGCTGTGAGGAACCTTGAGCTGGA 1420

Db 481 GAAGGGCGATCAATTGGAGCTTACGAGCTTGGCAAGTTGTTGAAGAACTGGAGCTGGA 540
QY 1421 GAAATATTATTGAACCTGCATTTGATTTGATGTTGATGTTGATGTTGATGATGATGATG 1480
DB 541 GAAATATTATTGAACCTGCATTTGATTTGATGTTGATGTTGATGATGATGATGATGAT 600
QY 1481 ATCAAGCTAATATCCGATGCTGTGAACATTCCTGTTATCGAAGACGCGTGCAGAGTC 1540
DB 601 ATAAAGCTGATATCTGATGCTGTCAACA-TCCTTGTATTGCAAGTAGTACGCTGCTGCT 659
QY 1541 GCTGATCACTTCTCGAAGTCTTTATGAACCAACGCGATCTGCTGCTTTCAGCTGGC 1600
DB 660 GTTGAACACTTCTCAGAGGT-ATTATGAACCAATGATGATGATGATGATGATGATG 718
QY 1601 ATTTTCCATCGCAAGAGGTTCCAATTAAGGCTGTTAAAGAGACACTTGTGTAAGGAAG 1658
DB 719 ATTTTCCATCGGAGAGGTGCCATTCAGTCTGTAAGACACACTTGTGTAAGGAAG 776

RESULT 3
BU046804
LOCUS
DEFINITION
PP_LEA0027M01f Peach developing fruit mesocarp Prunus persica cDNA
ACCESSION
BU046804
VERSION
BU046804.1 GI:22486881
KEYWORDS
EST.
SOURCE
Prunus persica (peach)
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
1 (bases 1 to 638)
Callahan,A., Palmer,M., Main,D., Wing,R. and Abbott,A.
Peach Model Genome for Rosaceae
Unpublished
Contact: Abbott, A.
Dept of Genetics and Biochemistry
Clemson University
122 Long Hall, Clemson University, Clemson, SC 29634, USA
Tel: 864 656 3060
Fax: 864 656 6879
Email: aalbert@clemson.edu
Total High Quality bases = 513
Seq primer: TAATACGACTCACTATAGG
High quality sequence stop: 638.

FEATURES
source
1..638
/organism="Prunus persica"
/mol_type="mRNA"
/cultivar="Loring"
/db_xref="taxon:3760"
/clone="PP_LEA0027M01f"
/tissue_type="Mesocarp"
/lab_host="E. coli"
/clone_lib="Peach developing fruit mesocarp"
/note="Vector: pBluescript II SK(-); Site 1: EcoRI; Site 2: XhoI; authority=Prunus persica L. Bateh; The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis go to
http://www.genome.clemson.edu/projects/peach. To order this clone go to http://www.genome.clemson.edu/orders"

BASE COUNT
188 a 92 c 174 g 182 t

ORIGIN
Query Match 21.9%; Score 428.4; DB 13; Length 638;
Best Local Similarity 79.7%; Pred. No. 4.3e-43;
Matches 504; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
QY 905 GATGAGTCACTTCTTAAACATTACTGTTTCCGGAGCTTCCCTTAGCGGATTCACCC 964

Db 1 GATGAGGTCACTTTTAAATATTACCGGTTTCGCGACTCCCTTTGGCGGACTTGCCA 60
Qy 965 ATGCTACAGGTCCTGCAACGGGATCTGAAACCTTTTGTGCGCATTAACCTGCGGGGT 1024
Db 61 ATGTTACAGGTAAGTACATACATACAGAAATGTTTGTACCATTAACAGTTGGAGGT 120
Qy 1025 GGCATCAGGATTTTACTGATGCAAAATGGAAGGTATTATTCTAGTCTAGAAAGTGGCTTCA 1084
Db 121 GGCATTAAGATTTTACAGATGCTATGCGAGAGTATTCTAGTTTGGAGTTGCTTCA 180
Qy 1085 GAGTATTTCAGATCGGGCGCGGATPAAGTTTTCGATCGGAAGTATGATGAGTTTACACTGCT 1144
Db 181 GAATATTTCAGATGCGGGCTGATAAGATTCCATTGGGAGTGATGAGTTTATGCTGCA 240
Qy 1145 GAGGAATATTAAACCGGAGTGAACACAGAGAAAGACATAGAGCAGATATCTACA 1204
Db 241 GAAGAAATATTAAAGACTGGAGTAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 300
Qy 1205 GTATATGTAACACGAGGAGTGTGTTAAGCATTCATCTCCGCGAGTTTACTTTGAGAAA 1264
Db 301 GTTTATGGAATCAGGCTGTGTTGTAAGCATTCATCTCCGAGTGTACCTTAAAT 360
Qy 1265 CCCGATGAAGTGAATTTAAAGCCATCAAAAGTAAAGCCATCCAGTCCAAAACGAGGAGAA 1324
Db 361 CCAGAGGATGTAGGTTCAAGACTATTAGGTTAAACAAACCCAGTCCAAAACGAGGAGAA 420
Qy 1325 TATGCTGTTATCAGTCACTGTTTAATGTTGGAGAGAGAGGAGACCCATCGGAGCTTAT 1384
Db 421 TTTGATGTTATCAGTGTACAGTTAGCGGTGGCGAGAGGCGGACCAATTTGAGGCTTAT 480
Qy 1385 GAATAGTCAAGGCTGTTGAGGACTTTGAGCTGGAGAAATATTATTGAAGTCACTGATGAT 1444
Db 481 GAGCTTGAAGAGGAGTTGAGGAGCTTGGAGCTGGAGAGTGAAGTGAAGTGAAGTGAAG 540
Qy 1445 TGTGATGTTCAAGGAAAGGATTCGATATAGATCTGATCAAGCTTAATATCCGATGCTGTG 1504
Db 541 TGTGATGTTCAAGGAAAGGATTTGATATAGATTTAAAGTGAAGTGAAGTGAAGTGAAG 600
Qy 1505 AACATTCCTGTTATCGCAAGCAGCGGTGCGAG 1536
Db 601 AGCATTCCTGTGATTGTAGTGTGCTGTG 632

RESULT 4
BM878913 668 bp mRNA linear EST 07-MAR-2002
LOCUS P9-002 Sweetpotato Ipomoea batatas cDNA similar to Glutamine
DEFINITION amidotransferase/cyclase, mRNA sequence.
BM878913
ACCESSION BM878913.1 GI:19246579
VERSION EST.
KEYWORDS Ipomoea batatas (sweet potato)
SOURCE Ipomoea batatas

ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Convolvulaceae; Ipomoea.
REFERENCE
Jarret,R.L., Dang,P., Seo,Y.W. and Jang,C.S.
TITLE Sweetpotato ESTs
JOURNAL Unpublished
COMMENT Contact: Jarret RL
Bldg 4457
USDA/ARS, Plant Genetic Resources
1109 Experiment Street, Griffin, GA 30223, USA
Tel: 770/228-7303
Fax: 770/228-7282
Email: bjarret@aes.griffin.peachnet.edu.

FEATURES
source
1..668
/organism="Ipomoea batatas"
/mol_type="mRNA"
/cultivar="Jewel"
/db_xref="taxon:4120"

/tissue type="Entire in vitro plantlet"
/clone lib="Sweetpotato"
BASE COUNT 193 a 100 c 186 g 189 t
ORIGIN
Query Match 21.8%; Score 425.6; DB 12; Length 668;
Best Local Similarity 78.3%; Pred. No. 9.1e-43;
Matches 523; Conservative 0; Mismatches 144; Indels 1; Gaps 1;
Qy 886 GTATTATTTAGACGGTCTGATGAGTCAAG-CTTCTTAAACATTACTGTTTCGGGACT 944
Db 1 GTATTACAAGGATGGAGCTGATGAGGTAGCCCTTTTGAATATTTACTGTTTCAGAGACT 60
Qy 945 TCCTCTACGCGATCTACCCATGCTACAGGTCTTGCAACGCGCATCTGAAACGTTTGTG 1004
Db 61 TCCTCTCGCGCATCTACCAATGTTCCAGGTATTGAGGTACATCAGAGATGTTTGTG 120
Qy 1005 TGCCATTAACTGTCGGGGGTGGCATCAGGGATTTTACTGATGCAATGGAAGTATTTAT 1064
Db 121 TGCCTTTAACTGTTGGAGCGGTATTAGAGACTTCACTGATGGAATGCGAGTACTACT 180
Qy 1065 CTAGTCTAGAAGTGGCTTCAGAGTATTTCAGATCGGGCCCGATAGGTTTCGATCGGAA 1124
Db 181 CTAGTTTGAAGTTCCTCTGAGTATTTCAGATCTGCTGCTGCAAGGTTCTCTATAGGAA 240
Qy 1125 GTGATCAGTTCACACTGCTGAGGAATATATTAAACCGGAGTGAAGACAGGAAAGAGCA 1184
Db 241 GTGATCAGTATATGCTGCTGAAGATATTGGAATCTGGAGTAAACCTGGAAGAGCA 300
Qy 1185 GCATAGAGCAGATATCTACAGATATATGGTAAACAGGAGTGGTTGTAAGCATTCCTC 1244
Db 301 GTCTGAGCAGATCTCTAGTGTGTTATGGAATCAGGAGTGGTTGTAAGTATTGATCCAA 360
Qy 1245 GCGAGTTCCTACTTGAGAAAACCGGATGAAGTGAATTTAAAGCCATCAAGTAAAGCCATC 1304
Db 361 GGAGGGTGTATTGGAAGGCCCAAGAGATGTAGAGTTCAAGACTGTCAAAGTGCAGAAC 420
Qy 1305 CAGGTCCTCAACGCTGAGGAATATGCTGCTGATCAGTGCATCTGTTAATGTTGGACGAGAAG 1364
Db 421 TAGTCCAAATGTTGAAGATATGCTTGGTACAGTGCACAGTGAATGTTGGACGAGAAG 480
Qy 1365 GGAGACCCATCGGAGCTTATGAAGTGAAGCTGTTGAGGAACCTTGAGAGCTGGAGAAA 1424
Db 481 GGAGACCAATTTGAGCATTTGAGCTTGCAAAAGCTGTTGAAGAAATTTGGGAGCTGGT 540
Qy 1425 TATTATTGAACCTGATGTTGATGCTCAAGGAAAAGATTCGATATAGATCTCATCA 1484
Db 541 TACTGTAAACCTGCATCGACTCGGATGGTCAAGGTAAGGATTTGATATAGACCTTATCA 600
Qy 1485 AGCTAATATCCGATGCTGTGAACATTCCTGTTATCGCAAGCAGCGGTGCGAGGAGTCGCTG 1544
Db 601 AGTAAATATCAGATGCTGTAAGCATCCAGTAATTCGAAGCAGTGGTGTGGAGCTGTTG 660
Qy 1545 ATCACTTC 1552
Db 661 AGCATTTTC 668

RESULT 5
BQ112714 727 bp mRNA linear EST 07-MAR-2003
LOCUS EST598290 mixed potato tissues Solanum tuberosum cDNA clone STMCI95
DEFINITION 5' end, mRNA sequence.
BQ112714
ACCESSION BQ112714
VERSION BQ112714.2 GI:21915593
KEYWORDS EST.
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanaceae; Solanum.
REFERENCE 1 (bases 1 to 727)
AUTHORS Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,

Restrepo, S., Griffiths, H., van der Hoeven, R., Teai, J. and Karamycheva, S.A.
 Generation of a set of potato cDNA clones for microarray analyses
 Unpublished
 On Apr 17, 2002 this sequence version replaced gi:20164676.

CONTACT: Robin Buell
 The Institute for Genomic Research
 9712 Medical Center Dr, Rockville, MD 20850, USA
 Email: potato-array@tigr.org
 This clone can be obtained from the University of Arizona Genomics
 Institute. Orders can be made through URL:
 http://genome.arizona.edu/orders/
 Seq primer: T3.

FEATURES

source

1. .727
 Location/Qualifiers
 /organism="Solanum tuberosum"
 /mol_type="mRNA"
 /cultivar="Kennebec or Binjite"
 /db_xref="taxon:4113"
 /clone="STMC195"
 /tissue_type="mixed tissues"
 /lab_host="SOLR"
 /clone_lib="mixed potato tissues"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Combination of untreated and Phytophthora infestans-treated libraries of stolons, leaves, leaflets, axillary buds of stem explants, petioles, germinating eyes tubers, or roots."
 191 a 160 c 215 t

BASE COUNT

ORIGIN

Query Match 17.2%; Score 335.8; DB 13; Length 727;
 Best Local Similarity 74.4%; Pred. No. 7.1e-32;
 Matches 437; Conservative 0; Mismatches 147; Indels 3; Gaps 1;

QY 55 CGCTTGGCTGGTGCAGCGGAGGATCTGTGGTGAATTTACTTGTATTCAGTTCGCTGGAAA 114
 |||||
 DB 141 CGCTTCGGCCACCAGCGCGGATGATCTGTGGTAACTGTACTTGTGATTGGTCTGGCAA 200
 |||||

QY 115 TGTAGGAGTGTAGGAGACCCATCCGACACTGGATTTGATTCATCAAGATGTGCAAAA 174
 |||||

DB 201 TGTGCGGAGTCTCAGGAATGCAATTAACATCTCGGTTTTGACATCAAGATGTCANAA 260
 |||||

QY 175 GCCAGAGGATATCTAAATGCTAAGCGCCTTACTTTCTGGCGTTGGGCGCTTTCGACC 234
 |||||

DB 261 ACCAGAGGACATTTTGAACCAAAACGACTCATTTTCTGGGTTGGTGTCTTTGCTCC 320
 |||||

QY 235 TCGAATGGATGTCTTATTCGTAAAGGGCTGGCTGAAGCACTCTGTACTTACATTCAGAA 294
 |||||

DB 321 TGCCATGGATGTCTAAACAAGAAGGATGGCTGAAGCACTTTGTCTTACATTCAGCA 380
 |||||

QY 295 TGATCGACCTTCTCGGTATATCGCTGGGATTCGACGACTCTTTGAGTCAAGTGAAGA 354
 |||||

DB 381 AGATCGCCCATCTCTAGGCAATTTGTCTGACGTCAGACTACTTTTGGTCAAGTGAAGA 440
 |||||

QY 355 AATGTGTCCAAATCAAGGCTCTGGCTTGATTCCTGGACGGGTTGGGCGTTTTCGAATCATC 414
 |||||

DB 441 AATGGACCACTAAAGGCTCTAGTTTGATTCCTGGAGTAGTTGGAGCGTTTTCGACTCTTC 500
 |||||

QY 415 CAATGGTTTAAAGGGTGCACATATGGATGGCATGCCCTTGGATATAAAGGAAGGGTCAGC 474
 |||||

DB 501 TAAAGGATCAGAGTACCTCACATTCGCTGGATGCACCTTGAATACGAAAAGATACTCA 560
 |||||

QY 475 AATTTTAGATGTGGGGAATCAACATGTGTATTTTGTTCATCTCATATCGACCCA---A 531
 |||||

DB 561 AATTTTGGATGACGTTTGGAAATTCACATGTCTATTTTCGTACATCTTATTCGGCAATGCC 620
 |||||

QY 532 TGCAGGAGCAACAAGAGTGGATTTTCATCTACATCGACGTATGGTGACGATTTTATTCG 591
 |||||

DB 621 GTCTGATGAATATAGAGGTGGATATCATCGACTTCGACGTCAGTACGGTACGATTTATAGC 680
 |||||

QY 592 ATCCATTCAGAGGGGAAATGTTTCATGTCAGTCCCAATTTTCATCCCGAGA 638
 |||||

Db 681 ATCCATTACAGAGGGGAAATGTCCATGCAGTTCAGTTTCATCTCTGAGA 727

RESULT 6
BG833389/cLOCUS
DEFINITIONACCESSION
VERSIONKEYWORDS
SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT


```

/dev_stage="1 to 3 days"
/lab_host="SOLR"
/clone_lib="potato stolon, Cornell University"
/notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; RNA was supplied by Christian Bachem & Beatriz
Horvath(Laboratory of Plant Breeding, Dept. of Plant
Sciences, Wageningen University, The Netherlands). Total
RNA was isolated from developing axillary buds of potato
nodal stem cuttings cultured on medium for the
introduction of tuber formation as described in Bachem et
al. (Plant Journal 1996). Tissue samples were taken of
stages corresponding to growing stolons and the early
stages of tuber formation."
BASE COUNT 147 a 88 c 154 g 144 t
ORIGIN
Query Match 16.7%; Score 327.2; DB 10; Length 533;
Best Local Similarity 75.3%; Pred. No. 9.6e-31;
Matches 404; Conservative 0; Mismatches 128; Indels 0; Gaps 0;
QY 854 AACCTTGGCAAGCCTGTGTGAACCTTGTGGCGCAGTATTATTAGACGGTGCTGATGAGGTC 913
Db 2 AACCTTGGCAAGCCCGTGGATCTGTGTGGACAGTATTACAAGGATGGGCGAGATGAGGTC 61
QY 914 AGCTTCTTAAACATTACTGTTTCCGGAGCTTCCCTCTAGGCGATCTACCCATGCTACAG 973
Db 62 AGCTTTTAAATATTACTGTTTCCGAGATTTCCCTTGGCGATCTGCCAATGTTGCAG 121
QY 974 GTCTTCGACGCCATCTGAAACGTTTTGTGCCATTAACTGTGGGGTGCCATCAGG 1033
Db 122 GTATTGAGGTATGATCAGAGAAATGTTTTGTGCCATTAACTGTGGAGTGATATTAGA 181
QY 1034 GATTTTACTGATCAAAATGGAAGTATTATCTAGTCTAGAAGTGGCTTCAGAGTATTTTC 1093
Db 182 GACTTCAAGATGCGAATGCGACATCTACTCTAGTTTGGAAATGCTTCAGAAATTTT 241
QY 1094 AGATCGGGCGCGATGAAGTTTCGATCGGAAGTGATGACGTTTACACTGCTGAGGAATAT 1153
Db 242 CGTTCTGGGCGACACAAAATTTCTATTGGAAGTGATGCTGTTTATGCTGGGAAGAATAC 301
QY 1154 ATTAAACCGGAGTGAAGACAGCAAGAGCAGCATAGACGAGATATCTACAGTATATGTT 1213
Db 302 TTAATAATCTGGAGTGAACCTGGAAGAGCAGTCTAGACGAGATCTCCGAGTTTATGGA 361
QY 1214 AACACGCGAGTGTGTGAACATTTGATCTCGCGAGTTTACTTTGAGAAACCCGATGAA 1273
Db 362 ATACGCGAGTGTGTGAACATTTGATCTCGAGGGGTGTGTGAAGACCCCAAGGCT 421
QY 1274 GTAGAAATTTAAAGCCCATCAAAGTAAGCCATCCAGGTCCAAACCGTGAGGAATATGCTGG 1333
Db 422 GTGGAATTCAGGACTGTGAAGTGAGAAACCCAGGTCCAAATCGAGAGATGATGATGG 481
QY 1334 TATCAGTGCACTGTTTAAATGTGACGAGAAAGGAGACCCCATCGAGGTTATG 1385
Db 482 TACCAGTGCAGGTGATGTTGGGAGAGAGGCGGCCCTATTGAGGCTTATG 533

RESULT 9
BF519112 560 bp mRNA linear EST 08-DEC-2000
LOCUS EST456572 DSIL Medicago truncatula cDNA clone pBSIL-19B12, mRNA
DEFINITION sequence.
ACCESSION BF519112
VERSION BF519112.1 GI:11607795
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 560)
AUTHORS Fedorova, M., Pierson, B.L., Samac, D.A., Vance, C.P., Gantt, G.S., Peng

```

```

H., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S.,
Holt, I.B. and Fraser, C.M.
ESTs from leaves of Medicago truncatula after inoculation with
Colletotrichum trifolii
Unpublished
Contact: Deborah A. Samac
Department of Plant Pathology
University of Minnesota
495 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108, USA
Tel: 612 625 1243
Fax: 651 649 5058
Email: debysopuccini.crl.umn.edu
University of Minnesota name: M276541e
TIGR sequence name: MTFCD06TK
More information is available at: http://chrysie.tamu.edu/medicago
Seq primer: SKmod (CTA GAA CTA gtg gat CC).
FEATURES
Location/Qualifiers
1..560
/organism="Medicago truncatula"
/mol_type="mRNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pDSIL-19B12"
/tissue_type="leaves infected with Colletotrichum
trifolii"
/dev_stage="cotyledons and primary leaves harvested 5 and
8 days after inoculation with Colletotrichum trifolii"
/lab_host="E. coli strain XL0LR"
/clone_lib="DSIL"
/notes="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA from
cotyledons and primary leaves harvested 5 and 8 days after
inoculation with Colletotrichum trifolii. The cDNA was
directionally ligated into the Uni-ZAP XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-ZAP phage using EX-Aassist
helper phage and propagated in XL0LR cells. Note: EST may
be of fungal origin."
BASE COUNT 170 a 96 c 129 g 165 t
ORIGIN
Query Match 16.2%; Score 317.2; DB 10; Length 560;
Best Local Similarity 76.7%; Pred. No. 1.5e-29;
Matches 388; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
QY 1173 CAGGAAAGAGCAGCAGTAGAGCAGATATCTACAGTATATGTTACACGAGCAGTGTGTTAA 1232
Db 1 CAGGAAAGACCAAGCTTAGAGCAGATATCAATAGTTTATGGGAATCAGGCGAGTGTGTTA 60
QY 1233 GCATTGATCCTCGCCGAGTTTACTTGAGAAACCCGATGAAGTATTAAGCCCATCA 1292
Db 61 GTATTGATCCCCGTATGGTTTACATAAACAAATGATGTGCAATTCAGACCATTA 120
QY 1293 AAGTAAGCCATCCAGGTCCTCAACCGGTGAGGAATATGCTCGGTATCAGTCACCTGTTAATG 1352
Db 121 AGTTTCAAGTCCAGTCCCAATGGAGAGATATGCTGTTATCAGTACAGTTAATG 180
QY 1353 GTGACGAGAGAGGAGACCCATCGGAGCTTATGAACCTAGCTAGGCTGTTGAGGAACCTTG 1412
Db 181 GGGGGCGAGAGGTCGGCAATTTGGTCTTATGAACCTGGCGAAAGCTGTTGAAGAACTTG 240
QY 1413 GAGCTGGAGAAATATTATTGAACCTGATGTTGATGCTCAAGGAAAGGATTCGATA 1472
Db 241 GTGCTGGTGAATACTATTAAATTTGATTTGATGTTGATGTTGATGTTGATGTTGATGTTG 300
QY 1473 TAGATCTGATCAAGCTAATATCCGATGCTGTGAACATTCCTCTGTTATCGCAAGCAGCGGTG 1532
Db 301 TAGCTTTAATTAAATGATCTCAATCGCTCAATATCCCTGTGTTGCAAGTAGTGGTG 360
QY 1533 CAGGAGTCCGCTCATCACTTCTCCGAGTCTTTTAATGAACCAACGATCTGCTGCCCTTG 1592
Db 361 CTGGCATTCTTAATCACTTTTCCGATGTTTCCACTAAACCAATGATGATACGCTGCACCTTG 420

```


/tissue_type="somatic embryos cultured on MSD 20"
 /lab_host="DH10B"
 /clone_lib="Gm-cl036"
 /notes="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; This
 cDNA library was constructed from mRNA isolated from
 somatic embryos (age ranging from 2 months to 9 months)
 cultured on MSD 20. The library was prepared using the
 Life Technologies pSuperScript cDNA library construction
 kit. Complementary DNA was synthesized from mRNA using a
 poly (dT) sequence with a NotI restriction site. SalI
 linkers adapters were ligated to the blunt-ended cDNA
 fragments followed by NotI digestion. The cDNA fragments
 were directionally cloned into the NotI-SalI restriction
 site of the pSPORT1 vector. The ligated cDNA fragments
 were transformed into E.coli Electromax DH10B host cells.
 This library was constructed in the laboratory of Dr. Lila
 Vodkin by Anu Khanna at the University of Illinois at
 Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"

BASE COUNT 179 a 91 c 149 g 172 t
 ORIGIN
 Query Match 15.8%; Score 308.2; DB 12; Length 591;
 Best Local Similarity 76.9%; Pred. No. 1.8e-28; Indels 0; Gaps 0;
 Matches 376; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
 QY 1190 GAGCAGATATCTACAGTATATGTTAACCAGCAGTGGTTGTAAGCATTTGATCTCGCGCA 1249
 DB 1 GAGCAGATATCAAGAGTTTATGGATCATCAGCAGTGGTTGTTAGTATCGATCCCCGTAGA 60
 QY 1250 GTTTACTTGGAAAAACCCGATGAAGTATGATTTTAAAGCCATCAAGTAAGCCATCCAGGT 1309
 DB 61 GTGTATGTAAGAGTATCTAATGATGTGCAATTTGAAGACCATAAGGGTTTCAAGTCCAGGT 120
 QY 1310 CCAACGGTCAGGAATATCTGTTATCAGTGCACCTGTTAATGTTGACAGCAGAGGAGA 1369
 DB 121 TCAATGAGAGGAATATGATGTTATCAATGTACAGTTAATGGGGACGAGAGGCGCG 180
 QY 1370 CCATCGAGCTTATGAATAGCTTAAGCTGTTGAGCAATCTTGAGCTGCGAGAAATATTA 1429
 DB 181 CCAATTGGTCTTATGAATAGCTAGCAAAAGCAGTTGAGAGCTTGGTCTGGTGAATACTA 240
 QY 1430 TTGAATGCAATGATTTGATGTTGTCAGGAAAGGATTCGATATAGATCTGTATCAAGCTA 1489
 DB 241 CTTAATTGCAATGATTCGAGCGGTCAAGGGAAGGATTTGATGATAGATTTAATTAAGTTG 300
 QY 1490 ATATCCATCTGTGAATCTCTGTTATCGCAGCAGCGGTGACAGAGTCTGATCAC 1549
 DB 301 ATATCAGATCTGTGAATCTCTGTTATCGCAAGTAGCGGTGCTGCTCTGAACAC 360
 QY 1550 TTCTCCGAAGTCTTTAATGAACCAACGATCTGCTCCCTTGAGCTGGCATTTTCCAT 1609
 DB 361 TTCTCTGAGGTGTTCTAATAAACAATGCAATCAGCAGCATTTGCTGCTGCAATTTTTCAC 420
 QY 1610 CGCAAGAGGTTCCTAATTAAGGCTGTTTAAAGAGCAGCTTTGTTGAAGGAGGATTTGAAGT 1669
 DB 421 AGGAAGAGGTGCTTATTCAGTCGGTAAAGAGCATTGTTTGAAGGAGGATAGAGTT 480
 QY 1670 AGATTGTA 1678
 DB 481 CGAATATGA 489

RESULT 12
 BQ869893
 LOCUS QGD7106.yg.ab1 QG_ABCDI1 lettuce salinas linear EST 15-AUG-2002
 DEFINITION QGD7106, mRNA sequence.
 ACCESSION BQ869893
 VERSION BQ869893.1 GI:22256201
 KEYWORDS EST.
 SOURCE Lactuca sativa
 ORGANISM Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
 Cichorieae; Lactuca.
 1 (bases 1 to 716)
 Lin.H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison
 P., Kolkman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,
 Church,S., Jackson,L. and Bradford,K.
 Lettuce and Sunflower ESTs from the Compositae Genome Project
 http://compgenomics.ucdavis.edu/
 Unpublished
 Contact: Alexander Kozik [R.W.Michelmoe]
 Department of Vegetable Crops, R.W.Michelmoe Lab
 University of California at Davis (UCD)
 Asmunsen Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@ucdavis.edu [michelmoe@vegmail.ucdavis.edu]
 singleton, see http://cgpdb.ucdavis.edu/ for details.
 Plate: QGD7 row: I column: 06.

FEATURES
 Location/Qualifiers
 1..716
 /organism="Lactuca sativa"
 /mol_type="mRNA"
 /cultivar="Salinas"
 /db_xref="taxon:4236"
 /clone="QGD7106"
 /lab_host="E.coli"
 /clone_lib="QG_ABCDI1 lettuce salinas"
 /note="Vector: pBRCNDSFIAB; The library was constructed
 from 10 different sources of RNA from a single genotype.
 Separate cDNAs were generated using primers that
 incorporated unique 5' and 3' tags to distinguish each
 source of RNA. cDNAs were then pooled, size-fractionated,
 directionally cloned into a custom medium-copy vector and
 transformations made with four size classes to minimize
 size bias. Details of each source of RNA and library
 construction can be obtained at http://cgpdb.ucdavis.edu/
 TAG_LTB=QG_ABCDI1 lettuce salinas
 TAG_TISSUE=chemical induction
 TAG_SEQ=GTGAGCGGG"

BASE COUNT 184 a 138 c 177 g 217 t
 ORIGIN

Query Match 15.4%; Score 301.4; DB 13; Length 716;
 Best Local Similarity 73.5%; Pred. No. 1.1e-27;
 Matches 399; Conservative 0; Mismatches 141; Indels 3; Gaps 1;
 QY 55 CGCTTTGGCTGTGTCAGCGGAGATTCTGTGGTGACTTTACTTTGATTACGGTCTGAAA 114
 DB 173 CGCTTCCGCTGTTGGAGGCGACGATTCTCGGTGACTCTCTTGATTATGGTCCGGTAA 232
 QY 115 TGTTAGAGTGTGAGGAACGCCATCCGCACACTTGGATTGATATCAAGATGTGCAAAA 174
 DB 233 TGTTCAAAGTATTAGAAATGCAATCCGGTATCTCGGTCTCGATATCAAGATGTGAAAC 292
 QY 175 GCCAGAGGATATCTTAATGCTAAGCGCTTACTTTCTCTGGGCTTGGGCCCTTTGCACC 234
 DB 293 ACCGGAGGACATTCTGAATGCAAGCGCCTTATTGTTCTGGTGTGGAGCATTTGCTGC 352
 QY 235 TGCAATGGATGTTCTTTATTTCGTAAAGGGCTGGCTGAAGCAGCTCTGTACTTTACATTGAGAA 294
 DB 353 AATGATGGACGCTCTTAAACAACAACGCGATGGCTGAAGCAGCTGTGCACCTTACATAGAA 412
 QY 295 TGATCGACCTTTCCTCGGATATATGCTCGGATTTGACGCTACTCTTTGAGTCAAGTGAAGA 354
 DB 413 TGATCGCCCATTTTATAGGATTTGTCTGCTCGACGTACTCTTTGAGTCAAGTGAAGA 472
 QY 355 AATGCTCCAAATTCAGGTCTTGGCTTGTATCTCGACCGGTGGCGGCTTTTGAATCATC 414
 DB 473 AAACGGGCCAGTGAGAGGTCTTTGGTTGTATTCCTGGGGTGGTGGACGTTTTCCTTT 532
 QY 415 CAATGGTTTAAAGGGTGCCACATATTGGATGGCATGCTTGGATATAAAGGAGGGTCAGC 474


```

RESULT 14
BE353528
LOCUS
DEFINITION
  BE353528 tomato flower buds 0-3 mm, Cornell University
  Lycopersicon esculentum cDNA clone cTOA1807, mRNA sequence.
ACCESSION
  BE353528
VERSION
  BE353528.1 GI:9291420
KEYWORDS
  EST.
SOURCE
  Lycopersicon esculentum (tomato)
  Lycopersicon esculentum
  Lycopersicon esculentum
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
  asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.
  1 (bases 1 to 569)
REFERENCE
  van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang
  ,F., Hansen,T., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W.,
  Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
  Generation of ESTs from tomato flower tissue, 0-3 mm buds
  Unpublished
  Contact: CUGI
  Clemson University Genomics Institute
  Clemson University
  100 Jordan Hall, Clemson, SC 29634, USA
  Email: http://www.genome.clemson.edu/orders/index.html
  5 prime sequence.
FEATURES
  source
    1..569
    /location/Qualifiers
    /organism="Lycopersicon esculentum"
    /mol_type="mRNA"
    /cultivar="TA496"
    /db_xref="taxon:4081"
    /clone="cTOA18J7"
    /tissue_type="flower"
    /dev_stage="0-3mm buds"
    /clone_lib="tomato flower buds 0-3 mm, Cornell University"
    /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
    XhoI; supplier: Tanksley; Flower buds and flowers were
    taken from greenhouse plants (4-8 wks old, TA496). They
    were immediately frozen in liquid nitrogen and then
    size-separated while remaining frozen."
  BASE COUNT
    156 a 113 c 136 g 164 t
  ORIGIN
    Query Match 15.3%; Score 298.4; DB 10; Length 569;
    Best Local Similarity 73.3%; Pred. No. 2.9e-27;
    Matches 396; Conservative 0; Mismatches 141; Indels 3; Gaps 1;

  QY 55 CGCTTGGCTGGTGCAGCGGAGATTCTGTGGTGACTTACTTGAATTACGGTCTGGAAA 114
  DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  30 CGCTTCGCCACCAAGTCCGATGATCTGTGGTAACCTACTTGAATTATGCTGCTGGCAA 89

  QY 115 TGTAGGAGTGTGAGGAACCCATCCGCACATTTGGATTGATATCAAGATGTCGAAA 174
  DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  90 TGTGCGGAGTGTGAGGAATCAATTAATATCTCGGTTTCGACATCAAGATGTCGAAA 149

  QY 175 GCCAGAGGATTTCTAAATGCTAAGCGCTTATCTTCTCGGGGTTGGGCGCTTTGCACC 234
  DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  150 ACCACAGGACATTTTGAAGCAAAACGACTCATTTTCTCGGGGTTGGTGTCTTGTCTCC 209

  QY 235 TGCATAGGATGTTCTTATTCGTAAAGGGCTGGCTGAAGCACTCTGTACTTACATTCAGAA 294
  DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  210 TGCCATGGATGTTCTTAAACAAAGGAATGGCTGAAGCACTTGTGCTTACATTCAGCA 269

  QY 295 TGATCGACCTTCTCTGGGTATATCCCTGGGATTCGAGTACTCTTTGAGTCAAGTGAAGA 354
  DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  270 AGATCGCCCATTCCTAGGCAATTTGTCTTGACTGCAGTACTCTTTTGGTCAAGTGAAGA 329

  QY 355 AAATGTGCAATTCAGGCTCTCGCTTCTGATTCTCGACGGGTTGGGCGTTTTCGAATCATC 414
  DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  330 AAATGGACAAGTAAAGGGTCTTGTTTGAATTCCTGGAGTAGTTGGACGTTTTCGACTTTC 389

  QY 415 CAATGGTTTAAAGGGTGCACATATTGGATGGCATGCCCTTGGATATAAAGGAAGGGTCAGC 474
  DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 390 TAAAGGATCAGAGTACCTCATTGGCTGGAAATGCATTGAAATAGCAAAAGATACTCA 449
QY 475 AATTTTAGATGATGTGGGAAATCAACATGTGTATTGTTTCTACTCATATCGAGCCA---A 531
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 450 AATTTTGGATGAGTTGGAAATTCACATGTCTTATTTCGTACATTTCTTATCGGCATGCC 509
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 532 TGCCGAGGACAAAGAGTGGATTTCATCTACATGACGATATGTGTGAGCATTTTATTGTC 591
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 510 GTCAGATGAAATAGAGAGTGGATATCATCGACCTGCAAGATATGGTGAGCATTTTCATAGC 569

RESULT 15
BG833390/c
LOCUS
DEFINITION
  BG833390 511 bp mRNA linear EST 23-MAY-2001
  951007E08.x2 951 - BMS tissue from Walbot Lab (GR) Zea mays cDNA,
  mRNA sequence.
ACCESSION
  BG833390
VERSION
  BG833390.1 GI:14191160
KEYWORDS
  EST.
SOURCE
  Zea mays
  Zea mays
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
  clade; Panicoideae; Andropogoneae; Zea.
  1 (bases 1 to 511)
REFERENCE
  Maize ESTs from various cDNA libraries sequenced at Stanford
  Walbot,V.
  University
  Unpublished
  Contact: Walbot V
  Department of Biological Sciences
  Stanford University
  855 California Ave, Palo Alto, CA 94304, USA
  Tel: 650 723 2227
  Fax: 650 725 8221
  Email: walbot@stanford.edu
  Plate: 951007 row: E column: 08.
  Location/Qualifiers
    1..511
    /organism="Zea mays"
    /mol_type="mRNA"
    /cultivar="BMS (Black Mexican Sweet)"
    /db_xref="taxon:4577"
    /tissue_type="suspension culture"
    /dev_stage="mixed logarithmic and stationary growth
    phases"
    /lab_host="DH10B"
    /clone_lib="951 - BMS tissue from Walbot Lab (GR)"
    /note="Vector: pUC19; Site 1: EcoRI; Site 2: EcoRI; The
    library was prepared by George Rudenko using poly (A)
    selected RNA and Universal Riboclone cDNA Synthesis System
    (Promega). cDNA was synthesized using both random and
    oligo(dT) primers in separate reactions and equipped with
    EcoRI adaptors. Library was size-fractionated on agarose
    gels (for insert size >400bp) and non-directionally cloned
    into EcoRI-digested pUC19 vector. Blue/white selection on
    carbenicillin-containing plates was used to recover
    positive clones."
  BASE COUNT
    138 a 144 c 84 g 145 t
  ORIGIN
    Query Match 14.8%; Score 289; DB 12; Length 511;
    Best Local Similarity 73.3%; Pred. No. 4.3e-26;
    Matches 370; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

  QY 1174 AGGAAAGACAGCATAGACAGATATCTACAGTATATGTAACCGAGGAGTGGTTGAAG 1233
  DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  511 AGGAAAGACAGCTTGGAGCAATCTCTAGAGTATATGCAATCAGGCTGTAGTTGTCAG 452

  QY 1234 CATTTGATCTCCCGAGTTTACTTGAGAAAACCCGATGAAGTAGAATTTAAAGCCATCAA 1293
  DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  451 TATTGATCTCGACGGGTATATGTCAAAGATGTGCCATTTTAAAACTGTAAG 392

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:20:02 ; Search time 46 Seconds
(without alignments)
1925.422 Million cell updates/sec

Title: US-09-831-233A-2
Perfect score: 2877
Sequence: 1 HEKELASTKPFQNGFRIRAL.....EVPKAVKHEHLKKEGIEVRL 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2877	100.0	558	21	AA1994224
2	2284.5	79.4	592	23	ABB93115
3	2160	75.1	593	21	AA194231
4	2157	75.0	593	21	AA194236
5	2157	75.0	593	21	AA194242
6	2154	74.9	593	21	AA194235
7	2154	74.9	593	21	AA194237
8	2154	74.9	593	21	AA194238
9	2154	74.9	593	21	AA194241

10	2154	74.9	593	21	AA194244	Cress mutant cyclo
11	2154	74.9	593	21	AA194246	Cress mutant cyclo
12	2152	74.8	593	21	AA194239	Cress mutant cyclo
13	2152	74.8	593	21	AA194240	Cress mutant cyclo
14	2152	74.8	593	21	AA194243	Cress mutant cyclo
15	2152	74.8	593	21	AA194245	Cress mutant cyclo
16	2151	74.8	593	21	AA194232	Cress mutant cyclo
17	2150	74.7	593	21	AA194233	Cress mutant cyclo
18	2150	74.7	593	21	AA194234	Cress mutant cyclo
19	1279	44.5	295	21	AA194226	Corn putative glut
20	630	21.9	141	21	AA194230	Soybean putative g
21	481.5	16.7	258	22	AA194237	C glutamicum prote
22	481.5	16.7	258	22	AA194238	Corynebacterium g
23	467.5	16.2	251	23	ABB48751	Listeria monocytog
24	439	15.3	256	23	ABP65662	Bifidobacterium lo
25	430.5	15.0	258	23	ABP39802	staphylococcus epi
26	428	14.9	255	24	ABP77948	N. gonorrhoeae ami
27	415.5	14.4	256	22	AAU53170	Propionibacterium
28	413.5	14.4	247	22	AA194239	S. epidermidis ope
29	409	14.2	208	23	ABB48753	Listeria monocytog
30	399.5	13.9	259	23	ABB54545	Lactococcus lactis
31	357	12.4	192	20	AA197360	An amidotransferas
32	357	12.4	192	20	AA197361	Amidotransferase d
33	355	12.3	192	23	ABP39807	Staphylococcus epi
34	339.5	11.8	222	22	AAU63592	Propionibacterium
35	339	11.8	226	24	ABP77949	N. gonorrhoeae ami
36	338	11.7	202	23	ABB54543	Lactococcus lactis
37	336.5	11.7	215	23	ABP66036	Bifidobacterium lo
38	332	11.5	211	22	AA194240	C glutamicum prote
39	332	11.5	211	22	AA194241	Corynebacterium g
40	286	9.9	114	23	ABP33635	Human ORF2608 prot
41	285	9.9	130	21	AA194225	Corn putative glut
42	284	9.9	158	22	AA194225	S. epidermidis ope
43	264	9.2	251	18	AA194229	Hsp protein invol
44	262	8.4	58	21	AA194229	Soybean putative g
45	215	7.5	127	22	AA194229	S. epidermidis ope

ALIGNMENTS

RESULT 1
AA194224
ID AA194224 standard; Protein; 558 AA.
XX
AC AA194224;
XX
DT 08-AUG-2000 (first entry)
XX
DE Impatiens balsamia glutamine amidotransferase.
XX
KW Glutamine amidotransferase; histidine biosynthesis; herbicide;
KW fungicide.
XX
OS Impatiens balsamia.
XX
PN WO200028053-A2.
XX
PD 18-MAY-2000.
XX
PF 04-NOV-1999; 99WO-US25950.
XX
PR 05-NOV-1998; 98US-0107275.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E I.
PA
XX
PI Allen SM, Huang LL, Falco SC, Rafalski AJ;
XX
DR WPI: 2000-376564/32.
DR N-PSDB; AA27325.
XX
PT Novel polynucleotides encoding plant glutamine amidotransferase
PT homologues, useful for producing transgenic plants and as probes or

PT primers -
 XX Claim 10; Page 41; 53pp; English.
 XX The present sequence is the Impatiens balsamia glutamine
 CC amidotransferase protein sequence. The protein forms a crucial stage
 CC in the histidine biosynthesis pathway. The gene sequence can be
 CC used to create transgenic plants which express different amounts of the
 CC protein, to identify loss of function mutants and to produce the
 CC protein in a host cell, for example a bacterium. The protein can be used
 CC to identify inhibitors which may be useful as fungicides and herbicides.
 XX
 XX Sequence 558 AA;
 SQ
 Query Match 100.0%; Score 2877; DB 21; Length 558;
 Best Local Similarity 100.0%; Pred. No. 1.5e-252;
 Matches 558; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 HEKELASTKQNGFRIRAAAGAGGDSVVTLLDYGAGNVRSVRNAITRTGFDIKDVQKPE 60
 DB 1 HEKELASTKQNGFRIRAAAGAGGDSVVTLLDYGAGNVRSVRNAITRTGFDIKDVQKPE 60
 QY 61 DIINAKELIPFGVGAPAPAMDVLIRKGLAEALCTYIQNDPFLGICLGLQLLPESSENG 120
 DB 61 DIINAKELIPFGVGAPAPAMDVLIRKGLAEALCTYIQNDPFLGICLGLQLLPESSENG 120
 QY 121 PIQGLGLIPGVGRFESSNGLRPHIGWHDLDIKESGAILDDVGNQHVYFVHSYRANAED 180
 DB 121 PIQGLGLIPGVGRFESSNGLRPHIGWHDLDIKESGAILDDVGNQHVYFVHSYRANAED 180
 QY 181 NKEWISSTCSYGDDFTASIQKGNHVAQVPHPEKSGVGLSILRRFLNADSFNNKROKPMN 240
 DB 181 NKEWISSTCSYGDDFTASIQKGNHVAQVPHPEKSGVGLSILRRFLNADSFNNKROKPMN 240
 QY 241 GKASKLAKRVIACLDVRANDNGBLVTWKGDQYDVRTEENEVRNLGKPVLAGQYYLDG 300
 DB 241 GKASKLAKRVIACLDVRANDNGBLVTWKGDQYDVRTEENEVRNLGKPVLAGQYYLDG 300
 QY 301 ADEVSFLNIITGPRDFPLGDLPMQLQVLRASENVFVPLTVGGGIRDFTDANGRYSSLEVA 360
 DB 301 ADEVSFLNIITGPRDFPLGDLPMQLQVLRASENVFVPLTVGGGIRDFTDANGRYSSLEVA 360
 QY 361 SEVFRSGADKVSIGSDAVYTAEEYIKGTGVTGKSSIEQISTVYGNQAVVVSIDPRVYLR 420
 DB 361 SEVFRSGADKVSIGSDAVYTAEEYIKGTGVTGKSSIEQISTVYGNQAVVVSIDPRVYLR 420
 QY 421 KPDEVEPKAIKVSHPGNGBEYAWYQCTVNGGREGRPFGAYELAKAVEELGAGEILLNCI 480
 DB 421 KPDEVEPKAIKVSHPGNGBEYAWYQCTVNGGREGRPFGAYELAKAVEELGAGEILLNCI 480
 QY 481 DCDGQGGKGFIDILKILSDAVNIPVIASSGAGVADHSEVFNETNASAAALAGIFHRKEV 540
 DB 481 DCDGQGGKGFIDILKILSDAVNIPVIASSGAGVADHSEVFNETNASAAALAGIFHRKEV 540
 QY 541 PIKAVKEHLKKEGIEVRL 558
 DB 541 PIKAVKEHLKKEGIEVRL 558
 RESULT 2
 ABB93115
 ID ABB93115 standard; Protein; 592 AA.
 AC
 AC ABB93115;
 XX
 XX 31-MAY-2002 (first entry)
 DE Herbicidally active polypeptide SEQ ID NO 2326.
 XX Herbicidal; plant; agriculture; herbicide.
 XX Arabidopsis thaliana.
 XX

PN WO200210210-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 28-AUG-2001; 2001WO-BF09892.
 XX
 PR 28-AUG-2001; 2001WO-BF09892.
 XX
 PA (FARB) BAYER AG.
 XX
 PI Tietjen K, Weidler M;
 XX
 DR WPI; 2002-369010/31.
 XX
 XX Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 XX
 XX Claim 5; SEQ ID NO 2326; 261pp + Sequence Listing; English.
 XX
 XX The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.
 XX
 SQ Sequence 592 AA;
 Query Match 79.4%; Score 2284.5; DB 23; Length 592;
 Best Local Similarity 80.7%; Pred. No. 1.2e-198;
 Matches 435; Conservative 56; Mismatches 45; Indels 3; Gaps 2;
 QY 21 AGAGGDSVVTLLDYGAGNVRSVRNAITRTGFDIKDVQKPEDILNAKRLIPFGVGAFAPAM 80
 DB 56 ASTSDSVVTLLDYGAGNVRSVRNAITRTGFDIKDVQKPEDILNAKRLIPFGVGAFAPAM 115
 QY 81 DVLIRKGLAEALCTYIQNDPFLGICLGLQLLPESSENGPIQGLGIPGVGRFESSNG 140
 DB 116 DVLNRTGMAEALCKYIENDRPFLGICLGLQLLPDSSENGPVKGLGIPGVGRFESSAG 175
 QY 141 LRVPHTGWAHDIKESGAILDDVGNQHVYFVHSYRANAEDNKEWISSTCSYGDDFTASI 199
 DB 176 IRVPHIGWALQVKGSEILDDVGNRHVYFVHSYRAIPSDENKDWISSTCSYGESFSSI 235
 QY 200 QKGNHVAQVPHPEKSGVGLSILRRFLNADSFNNKROKPMNGKASKLAKRVIACLDVRAN 259
 DB 236 RRGNVHVAQVPHPEKSGVGLSILRRFLNADSFNNKROKPMNGKASKLAKRVIACLDVRAN 293
 QY 260 DNGDLVVTKGDDQYDVRTEENEVRNLGKPVLAGQYYLDGADDEVSLNITGPRDFPLGD 319
 DB 294 DKGDLVVTKGDDQYDVRTEENEVRNLGKPVLAGQYYLDGADDEVSLNITGPRDFPLGD 353
 QY 320 LPMQLQVLRASENVFVPLTVGGGIRDFTDANGRYSSLEVASSEYFSSGADKVSIGSDAVY 379
 DB 354 LPMQLQVLRASENVFVPLTVGGGIRDFTDANGRYSSLEVASSEYFSSGADKVSIGSDAVY 413
 QY 380 TAEYIKGTGVTGKSSIEQISTVYGNQAVVVSIDPRVYLRKPDEVEPKAIKVSHPGNG 439
 DB 414 AAEFIKSGVTKGSSLEQISRYGNQAVVVSIDPRVYLRKPDEVEPKAIKVSHPGNG 473
 QY 440 EEWYQCTVNGGREGRPFGAYELAKAVEELGAGEILLNCIDCDGQGGKGFIDILKILSD 499
 DB 474 EEWYQCTVNGGREGRPFGAYELAKAVEELGAGEILLNCIDCDGQGGKGFIDILKILSD 533
 QY 500 ANNIPVIASSGAGVADHSEVFNETNASAAALAGIFHRKEVPIKAVKEHLKKEGIEVRL 558
 DB 534 SVGIPVIASSGAGTDPHFSEVFEKTNASAAALAGIFHRKEVPIKAVKEHLKKEGIEVRL 592


```
RESULT 3
AA94231
ID AAY94231 standard; Protein; 593 AA.
XX
AC AAY94231;
XX
DT 08-AUG-2000 (first entry)
XX
DE Arabidopsis thaliana HisH enzyme.
XX
KW Glutamine amidotransferase; histidine biosynthesis; cyclase;
KW mouse ear cress; site-directed mutagenesis.
XX
OS Arabidopsis thaliana.
XX
PN WO200028053-A2.
XX
PD 18-MAY-2000.
XX
PF 04-NOV-1999; 99WO-US25950.
XX
PR 05-NOV-1998; 98US-0107275.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Huang LL, Falco SC, Rafalski AJ;
XX
DR WPI; 2000-376564/32.
XX
PT Novel polynucleotides encoding plant glutamine amidotransferase
PT homologues, useful for producing transgenic plants and as probes or
PT primers -
XX
PS Example 9; Page 50-52; 53pp; English.
XX
CC The present sequence is the mouse ear cress HisH enzyme, which is
CC involved in histidine biosynthesis. It was used in site-directed
CC mutagenesis to find the active sites within the enzyme, which is made up
CC of a cyclase and a glutamine amidotransferase. This is useful for the
CC identification of herbicides and fungicides which inhibit the enzyme.
SQ Sequence 593 AA;
Query Match 75.1%; Score 2160; DB 21; Length 593;
Best Local Similarity 76.7%; Pred. No. 2.5e-187;
Matches 414; Conservative 60; Mismatches 62; Indels 4; Gaps 3;
QY 21 AGAGDSVVTLLDYGAGNVRNRAIRTLGFDIKDVQKPEDILNAKRLIFPGVGAFAPAM 80
DB 56 ASSTSDSVVTLLDYGAGNVRNRAIRTLGFDIKDVQKPEDILNAKRLIFPGVGAFAPAM 115
QY 81 DVLIRKGLAALCTYQNDPFLGICLGLQLPESSENGPIQGLGILPGVRGPRESSNG 140
DB 116 DVLNRTGMAELCKYIENDRPFLGICLGLQLPESSENGPIQGLGILPGVRGPRESSNG 175
QY 141 LRVPHIGMHALDIKEGAILDDVGNHVVYFVHVSRA-NAEDNKEMISSTCSYGDDEFTASI 199
DB 176 IRVPHIGMNAQVKQSEIILDDVGNHVVYFVHVSRAIPSDENKDWISSTCSYGESFISSI 235
QY 200 QKGNVHAVQHPKSGVGLSILRRFLNADSFNNKQKPMNGKASKAKLAKVIACLDVRAN 259
DB 236 RRGNNHAVQHPKSGVGLSVLRRFLHPLK--PATQKPMEGKASKAKLAKVIACLDVRTN 293
QY 260 DNGDLVVTGQDVQDVRERTENEVRNLGKPVLAGQYVLDGAEVFSPLNITGFRDPLGD 319
DB 294 DKGDLVVTGQDVQDVRQSENEVRNLGKPVLAGQYVLDGAEVFSPLNITGFRDPLGD 353
QY 320 LPMIQLVLRASENVFPLTVGGGIRDTDANGRYSSLEVASVYERSGADKVISGDVAVY 379
DB 354 LPMIQLVLRQTSKNVFPVPLTVGGGIRDTDASGRYSSLEVAAEYFRSGADKMSGDAVF 413
QY 380 TABEYIKTKGSKSIEQISTVYGNQAVVVSIDPRRVYLRKPDEVEFKAIVKSHFGPNG 439
```

```
DB 414 AAEEFIKSGVTKGSKSLEQISRVYGNQAVVVSIDPRRVYVNHPPDDVPYVIRVTNPGNG 473
QY 440 EYAWYQCTVNGRGRRPIGAYELAKAVEELGAGEILLNCIDCGGKGKGFDDILKLIISD 499
DB 474 EYAWYQCTVSGGQGRPIGAFELAKAVEELGAGEILLNCIDCGGKGKGFDDILKLIISD 533
QY 500 AVNIPVIASSGAGVADHFSEVFNETNASAAAGIF-HRKEVPIKAVKSHLLKKEGIEVRL 558
DB 534 SVGIPVIASSGAGTDPHFSEVFEEDKRICRACCRHPPPERGYQSQSVKSHLQEEIEVRI 593

RESULT 4
AA94236
ID AAY94236 standard; Protein; 593 AA.
XX
AC AAY94236;
XX
DT 08-AUG-2000 (first entry)
XX
DE Cress mutant glutamine amidotransferase, E248Q substitution.
XX
KW Glutamine amidotransferase; histidine biosynthesis; cyclase;
KW mouse ear cress; site-directed mutagenesis; mutein; substitution;
KW mutant.
XX
OS Arabidopsis thaliana.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 248 /note= "wild-type Glu substituted by Gln"
XX
PN WO200028053-A2.
XX
PD 18-MAY-2000.
XX
PF 04-NOV-1999; 99WO-US25950.
XX
PR 05-NOV-1998; 98US-0107275.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Huang LL, Falco SC, Rafalski AJ;
XX
DR WPI; 2000-376564/32.
XX
PT Novel polynucleotides encoding plant glutamine amidotransferase
PT homologues, useful for producing transgenic plants and as probes or
PT primers -
XX
PS Example 9; Page -; 53pp; English.
XX
CC The present sequence is a mutant of the mouse ear cress HisH enzyme,
CC which is involved in histidine biosynthesis. It was produced by
CC site-directed mutagenesis, which was carried out to find the active
CC sites within the enzyme, which is made up of a cyclase and a glutamine
CC amidotransferase. The mutation in this case is in the glutamine
CC amidotransferase domain. Determining the active sites is useful for the
CC identification of herbicides and fungicides which inhibit the enzyme.
CC This residue was shown to form part of the active site.
CC Note: The present sequence is not shown in the specification but is
CC derived from the Arabidopsis thaliana wild-type His HF sequence given in
CC the sequence listing (see AAY94231).
XX
SQ Sequence 593 AA;
Query Match 75.0%; Score 2157; DB 21; Length 593;
Best Local Similarity 76.5%; Pred. No. 4.7e-187;
Matches 413; Conservative 61; Mismatches 62; Indels 4; Gaps 3;
QY 21 AGAGDSVVTLLDYGAGNVRNRAIRTLGFDIKDVQKPEDILNAKRLIFPGVGAFAPAM 80
```

Db 56 ASSTDSVVTLLDYGAGNVRISRNALRHIGFSIKDVQTPGDIILNADRLIPPGVGPAPAM 115
 Qy 81 DVLIRKGLAALCTYIONDRPFLGICLGLQLLPESSEENGPIQGLGLIPGRVGRFPSSNG 140
 Db 116 DVLNRTGMALCKYIENDRPFLGICLGLQLLPDSSEONGPVKGLGVPIGVGRFPDASAG 175
 Qy 141 LRVPHICGHALDIKEGSAIILDDVGNQHVYFVHSYRA-NAEDNKESWISSTCSYGGDDFTASI 199
 Db 176 IRVPHIGWALQVGKSEILDDVGNRHVYFVHSYRAIPSDENKDWISSTCNYESFTSSI 235
 Qy 200 QKGNVHAVQHPKSGVGLSILRRFLNADSFNNKQKPMNGKASKLAKRVIACLDVRAN 259
 Db 236 RRGNVHAVQHPKSGEVGLSVLRFLHPKL--PATQKPMEGKASKLAKRVIACLDVRTN 293
 Qy 260 DNGDLVVTGQDYDVRERTEENVRNLGKPVELAGQYILDGAEVSEFLNITGFRDFPLGD 319
 Db 294 DKGDLVVTGQDYDVRERTEENVRNLGKPVELAGQYILDGAEVSEFLNITGFRDFPLGD 353
 Qy 320 LPMQLVLRASENVFVPLTVGGGIRDTDANGRYYSLEVASEYFRSGADKVSIGSDAVY 379
 Db 354 LPMQLVLRQTSKNVFPVPLTVGGGIRDTDANGRYYSLEVASEYFRSGADKVSIGSDAVF 413
 Qy 380 TABEYIKGTGKSSIEQISTYVGNQAVVVSIDPRVYLKRPDEVEFKAIVKSHPGPNG 439
 Db 414 AAEFIKSGVTKGSSLEQISRYVGNQAVVVSIDPRVYVNHDPDVFPYKIVRTNPGNG 473
 Qy 440 EYAWYQCTVNGRGREGPIGAYELAKAVEELGAGEILLNCIDCGQKGFIDILKLISD 499
 Db 474 EYAWYQCTVSGQGEGRPIGAFELAKAVEELGAGEILLNCIDCGQKGFIDILKLISD 533
 Qy 500 AVNIPVTASSGAGVADHPSEVFNETNASALAAGIF-HRKEVPKAVKEHLKEGIEVRL 558
 Db 534 SVGIPVIASSGAGTPDHPSEVFEEDEKRICRACCRHPPPERGYQSQSVKEHLQERIEVRI 593

RESULT 5

AA94242
 ID AA94242 standard; Protein; 593 AA.
 AC AA94242;
 XX
 XX 08-AUG-2000 (first entry)
 XX Cress mutant cyclase, S409A substitution.
 XX Glutamine amidotransferase; histidine biosynthesis; cyclase;
 KW mouse ear cress; site-directed mutagenesis; muten; substitution;
 KW mutant.
 XX Arabidopsis thaliana.
 OS Synthetic.
 FT Key Location/Qualifiers
 FT Misc-difference 409 /note= "Wild-type Ser substituted by Ala"
 FT
 XX WO200028053-A2.
 XX
 XX 18-MAY-2000.
 XX
 XX 04-NOV-1999; 99WO-US25950.
 XX
 XX 05-NOV-1998; 98US-0107275.
 XX
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 XX Allen SM, Huang LL, Falco SC, Rafalski AJ;
 XX
 XX WPI; 2000-376564/32.
 XX
 XX Novel polynucleotides encoding plant glutamine amidotransferase
 PT homologues, useful for producing transgenic plants and as probes or
 PT primers -

XX
 PS
 XX Example 9; Page -, 53pp; English.
 CC The present sequence is a mutant of the mouse ear cress HisHFP enzyme,
 CC which is involved in histidine biosynthesis. It was produced by
 CC site-directed mutagenesis, which was carried out to find the active
 CC sites within the enzyme, which is made up of a cyclase and a glutamine
 CC amidotransferase. The mutation in this case is in the cyclase
 CC domain. Determining the active sites is useful for the identification of
 CC herbicides and fungicides which inhibit the enzyme. This residue is
 CC not expected to be within the active site.
 CC Note: The present sequence is not shown in the specification but is
 CC derived from the Arabidopsis thaliana wild-type His HF sequence given in
 CC the sequence listing (see AA94231).

XX Sequence 593 AA;

Query Match 75.0%; Score 2157; DB 21; Length 593;
 Best Local Similarity 76.5%; Pred. No. 4.7e-187;
 Matches 413; Conservative 61; Mismatches 62; Indels 4; Gaps 3;

Qy 21 AGAGGSVVTLLDYGAGNVRISRNALRHIGFSIKDVQTPGDIILNADRLIPPGVGPAPAM 80
 Db 56 ASSTDSVVTLLDYGAGNVRISRNALRHIGFSIKDVQTPGDIILNADRLIPPGVGPAPAM 115
 Qy 81 DVLIRKGLAALCTYIONDRPFLGICLGLQLLPESSEENGPIQGLGLIPGRVGRFPSSNG 140
 Db 116 DVLNRTGMALCKYIENDRPFLGICLGLQLLPDSSEONGPVKGLGVPIGVGRFPDASAG 175
 Qy 141 LRVPHICGHALDIKEGSAIILDDVGNQHVYFVHSYRA-NAEDNKESWISSTCSYGGDDFTASI 199
 Db 176 IRVPHIGWALQVGKSEILDDVGNRHVYFVHSYRAIPSDENKDWISSTCNYESFTSSI 235
 Qy 200 QKGNVHAVQHPKSGVGLSILRRFLNADSFNNKQKPMNGKASKLAKRVIACLDVRAN 259
 Db 236 RRGNVHAVQHPKSGEVGLSVLRFLHPKL--PATQKPMEGKASKLAKRVIACLDVRTN 293
 Qy 260 DNGDLVVTGQDYDVRERTEENVRNLGKPVELAGQYILDGAEVSEFLNITGFRDFPLGD 319
 Db 294 DKGDLVVTGQDYDVRERTEENVRNLGKPVELAGQYILDGAEVSEFLNITGFRDFPLGD 353
 Qy 320 LPMQLVLRASENVFVPLTVGGGIRDTDANGRYYSLEVASEYFRSGADKVSIGSDAVY 379
 Db 354 LPMQLVLRQTSKNVFPVPLTVGGGIRDTDANGRYYSLEVASEYFRSGADKVSIGSDAVF 413
 Qy 380 TABEYIKGTGKSSIEQISTYVGNQAVVVSIDPRVYLKRPDEVEFKAIVKSHPGPNG 439
 Db 414 AAEFIKSGVTKGSSLEQISRYVGNQAVVVSIDPRVYVNHDPDVFPYKIVRTNPGNG 473
 Qy 440 EYAWYQCTVNGRGREGPIGAYELAKAVEELGAGEILLNCIDCGQKGFIDILKLISD 499
 Db 474 EYAWYQCTVSGQGEGRPIGAFELAKAVEELGAGEILLNCIDCGQKGFIDILKLISD 533
 Qy 500 AVNIPVIASSGAGVADHPSEVFNETNASALAAGIF-HRKEVPKAVKEHLKEGIEVRL 558
 Db 534 SVGIPVIASSGAGTPDHPSEVFEEDEKRICRACCRHPPPERGYQSQSVKEHLQERIEVRI 593

RESULT 6

AA94235
 ID AA94235 standard; Protein; 593 AA.
 XX
 XX AA94235;
 XX
 XX 08-AUG-2000 (first entry)
 XX Cress mutant glutamine amidotransferase, E248A substitution.
 XX Glutamine amidotransferase; histidine biosynthesis; cyclase;
 KW mouse ear cress; site-directed mutagenesis; muten; substitution;
 KW mutant.
 XX Arabidopsis thaliana.
 OS

Db 116 DVLNRTGMAELCKYIENDRPFLGICLGLQLLFDSSEQNGPVKGLGVPIGVGRFDASAG 175
Qy 141 LRVPHIGWALDIKESGAILDDVGNQVYFVHSYRA-NAEDNKWISSTCSYGGDDFTASI 199
Db 176 IRVPHIGWALQVKGKSEILDDVGNRHVYFVHSYRAIPSDENKDWISSTCNGESPTSSI 235
Qy 200 QKGNVHAVQHPKESGKGLSILRRFLNADSFNNKROKPMNGKASKLAKRVIACLDVRAN 259
Db 236 RRGNVHAVQHPKESGKGLSVLRFLHPKL--PATQKPMGKASKLAKRVIACLDVRTN 293
Qy 260 DNGDLVVTGQDYDVRTERTEENVRNLGKPVLAGQYLLDCADEVSLNITGFRDPLD 319
Db 294 DKGLVVTAGQDYDVRQSENEVRNLGKPVLAGQYLLDCADEISFLNITGFRDPLD 353
Qy 320 LPMQLVLRASENVFVPLTVGGGIRDTDANGRYSSLEVASSEYFRSGADKVSIGSDAVY 379
Db 354 LPMQLVLRQTSKNVFPVPLTVGGGIRDTDANGRYSSLEVASSEYFRSGADKVSIGSDAVF 413
Qy 380 TAEYIKTGKTKSSIEQISTVYGNQAVVSDPRRVYLRKPDEVEFKAIKVSHPGNG 439
Db 414 AAEFIKSGVTKGSSLEQISRVYGNQAVVSDPRRVYVNHDDVPYKVRVTNPGNG 473
Qy 440 EEWYQCTVNGGREGPIGAYELAKAVEELGAGEILLNCIDCDGQKGFIDILIKLISD 499
Db 474 EEWYQCTVSGGQEGRPAGAFELAKAVEELGAGEILLNCIDCDGQKGFIDILVKLISD 533
Qy 500 AVNIPVIASSGAGVADHFSEVFNETHASALAAGIP-HRKEVPKAVEKHELLKEGIEVRL 558
Db 534 SVGPVVIASSGAGTPDHFSEVFEEDKRICRACCRHPPFPERGYQSQVKEHLOEBIEVRI 593

RESULT 8

AA94238
ID AAY94238 standard; Protein; 593 AA.
XX
AC AAY94238;
XX
DT 08-AUG-2000 (first entry)
XX
DE Cress mutant cyclase, D334A substitution.
XX
KW Glutamine amidotransferase; histidine biosynthesis; cyclase;
KW mouse ear cress; site-directed mutagenesis; mutein; substitution;
KW mutant.
XX
OS Arabidopsis thaliana.
XX
FH Synthetic.
XX
FT Key Location/Qualifiers
FT Misc-difference 334 /note= "Wild-type Asp is substituted by Ala"
XX
PN WO200028053-A2.
XX
PD 18-MAY-2000.
XX
PF 04-NOV-1999; 99WO-US25950.
XX
PR 05-NOV-1998; 98US-0107275.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Huang LL, Falco SC, Rafaleki AJ;
XX
DR WPI; 2000-376564/32.
XX
PT Novel polynucleotides encoding plant glutamine amidotransferase
PT homologues, useful for producing transgenic plants and as probes or
PT primers -
XX
PS Example 9; Page -; 53pp; English.
XX
CC The present sequence is a mutant of the mouse ear cress HisH enzyme,

CC which is involved in histidine biosynthesis. It was produced by
CC site-directed mutagenesis, which was carried out to find the active
CC sites within the enzyme, which is made up of a cyclase and a glutamine
CC amidotransferase. The mutation in this case is in the cyclase
CC domain. Determining the active sites is useful for the identification of
CC herbicides and fungicides which inhibit the enzyme. This residue is
CC not expected to be within the active site.
CC Note: The present sequence is not shown in the specification but is
CC derived from the Arabidopsis thaliana wild-type HisH sequence given in
CC the sequence listing (see AAY94231). D334A. The residue in the
CC Note: The mutant is stated as being Gly.
CC wild-type is stated as being Gly.

SQ Sequence 593 AA;

Query Match 74.9%; Score 2154; DB 21; Length 593;
Best Local Similarity 76.5%; Pred. No. 8.8e-187;
Matches 413; Conservative 60; Mismatches 63; Indels 4; Gaps 3;

Qy 21 AGAGGDSVVTLLDYGAGNVRSVRNARTLGFDIKDVKPEDILNAKRLIFPGVGAFAPAM 80
Db 56 ASTSDSVVTLLDYGAGNVRSIRNARHLGFSIKDVQTFDILNADRLIFPGVGAFAPAM 115
Qy 81 DVLIRKGLAEALCTYIQNDRPFLGICLGLQLLFESSENGPIQGLGIPRVGRFPSSNG 140
Db 116 DVLNRTGMAELCKYIENDRPFLGICLGLQLLFDSSEQNGPVKGLGVPIGVGRFDASAG 175
Qy 141 LRVPHIGWALDIKESGAILDDVGNQVYFVHSYRA-NAEDNKWISSTCSYGGDDFTASI 199
Db 176 IRVPHIGWALQVKGKSEILDDVGNRHVYFVHSYRAIPSDENKDWISSTCNGESPTSSI 235
Qy 200 QKGNVHAVQHPKESGKGLSILRRFLNADSFNNKROKPMNGKASKLAKRVIACLDVRAN 259
Db 236 RRGNVHAVQHPKESGKGLSVLRFLHPKL--PATQKPMGKASKLAKRVIACLDVRTN 293
Qy 260 DNGDLVVTGQDYDVRTERTEENVRNLGKPVLAGQYLLDCADEVSLNITGFRDPLD 319
Db 294 DKGLVVTAGQDYDVRQSENEVRNLGKPVLAGQYLLDCADEISFLNITGFRDPLD 353
Qy 320 LPMQLVLRASENVFVPLTVGGGIRDTDANGRYSSLEVASSEYFRSGADKVSIGSDAVY 379
Db 354 LPMQLVLRQTSKNVFPVPLTVGGGIRDTDANGRYSSLEVASSEYFRSGADKVSIGSDAVF 413
Qy 380 TAEYIKTGKTKSSIEQISTVYGNQAVVSDPRRVYLRKPDEVEFKAIKVSHPGNG 439
Db 414 AAEFIKSGVTKGSSLEQISRVYGNQAVVSDPRRVYVNHDDVPYKVRVTNPGNG 473
Qy 440 EEWYQCTVNGGREGPIGAYELAKAVEELGAGEILLNCIDCDGQKGFIDILIKLISD 499
Db 474 EEWYQCTVSGGQEGRPAGAFELAKAVEELGAGEILLNCIDCDGQKGFIDILVKLISD 533
Qy 500 AVNIPVIASSGAGVADHFSEVFNETHASALAAGIP-HRKEVPKAVEKHELLKEGIEVRL 558
Db 534 SVGPVVIASSGAGTPDHFSEVFEEDKRICRACCRHPPFPERGYQSQVKEHLOEBIEVRI 593

RESULT 9

AA94241
ID AAY94241 standard; Protein; 593 AA.
XX
AC AAY94241;
XX
DT 08-AUG-2000 (first entry)
XX
DE Cress mutant cyclase, K404A substitution.
XX
KW Glutamine amidotransferase; histidine biosynthesis; cyclase;
KW mouse ear cress; site-directed mutagenesis; mutein; substitution;
KW mutant.
XX
OS Arabidopsis thaliana.
OS Synthetic.
XX

Key Location/Qualifiers
Misc-difference 404 /note= "Wild-type Lys is substituted by Ala"
W0200028053-A2.
18-MAY-2000.
04-NOV-1999; 99WO-US25950.
05-NOV-1998; 98US-0107275.
(DUPO) DU PONT DE NEMOURS & CO E I.
Allen SM, Huang LL, Falco SC, Rafaleki AJ;
WPI; 2000-376564/32.
Novel polynucleotides encoding plant glutamine amidotransferase
PT homologues, useful for producing transgenic plants and as probes or
primers
Example 9; Page -: 53pp; English.
The present sequence is a mutant of the mouse ear cress HisHf enzyme,
which is involved in histidine biosynthesis. It was produced by
site-directed mutagenesis, which was carried out to find the active
sites within the enzyme, which is made up of a cyclase and a glutamine
amidotransferase. The mutation in this case is in the cyclase
domain. Determining the active sites is useful for the identification of
herbicides and fungicides which inhibit the enzyme. This residue was
shown to form part of the active site.
Note: The present sequence is not shown in the specification but is
derived from the Arabidopsis thaliana wild-type His HF sequence given in
the sequence listing (see AAY94231).

Query Match 74.9%; Score 2154; DB 21; Length 593;
Best Local Similarity 76.5%; Pred. No. 8.8e-187;
Matches 413; Conservative 60; Mismatches 63; Indels 4; Gaps 3;

QY 21 AGAGDSVVTLLDYGAGNVRNNAIRTLGFDIKDVQKPEDILNAKRLIFPGVGAFAPAM 80
DB 56 ASSTSDSVTLLDYGAGNVRNNAIRTLGFDIKDVQKPEDILNAKRLIFPGVGAFAPAM 115
QY 81 DVLRKGLAELCTYQNDPRPFLGICGLQLLPESSENGPIQGLIPGRVRFSSNG 140
DB 116 DVLRKGLAELCTYQNDPRPFLGICGLQLLPESSENGPIQGLIPGRVRFSSNG 175
QY 141 LRVPHIGWALDIKEGSAILLDVGNOHVYFVHSYRA-NAEDNKEWISSTCSYGGDFTASI 199
DB 176 IRVPHIGWALDIKEGSAILLDVGNOHVYFVHSYRA-NAEDNKEWISSTCSYGGDFTASI 235
QY 200 QKGNHVAQPHPKSGVGLSILRRFLNADSFNNKQKPMNGKASKLAKRVIACLDVRAN 259
DB 236 RRGNVHVAQPHPKSGVGLSILRRFLNADSFNNKQKPMNGKASKLAKRVIACLDVRAN 293
QY 260 DNGDLVTKGQYDVRTERTENEVRNLGKPVLAGQYLLDGADEVSPNLTGFRDPFLGD 319
DB 294 DNGDLVTKGQYDVRTERTENEVRNLGKPVLAGQYLLDGADEVSPNLTGFRDPFLGD 353
QY 320 LPMLOLORASENVVPLTVGGIRPTDANGRYYSLEVASVRSKADKVSIGSDAVY 379
DB 354 LPMLOLORASENVVPLTVGGIRPTDANGRYYSLEVASVRSKADKVSIGSDAVY 413
QY 380 TABEYIKTGKTSIEQISTYGVNOAVVVSIDPRRVYLRKDEVEFKAIVKSHHPNG 439
DB 414 AAEPIKSGVTKGKSLEQISRVYGVNOAVVVSIDPRRVYLRKDEVEFKAIVKSHHPNG 473
QY 440 EYAWYQCTVNGREGRPIGAYELAKAVEELGAGEILLNCIDCDGQKGFIDILIKLISD 499
DB 474 EYAWYQCTVNGREGRPIGAYELAKAVEELGAGEILLNCIDCDGQKGFIDILIKLISD 533

QY 500 AVNIPVIASSGAGVADHFSEVFNETNASAAAGIF-HRKEVPKAVKHLKEGIEVRL 558
DB 534 SVGIPVIASSGAGTDPHFSEVFNETNASAAAGIF-HRKEVPKAVKHLKEGIEVRL 593

RESULT 10

AAY94244

XX AAY94244 standard; Protein; 593 AA.

XX AAY94244;

XX 08-AUG-2000 (first entry)

XX Cress mutant cyclase, E508A substitution.

XX Glutamine amidotransferase; histidine biosynthesis; cyclase;
XX mouse ear cress; site-directed mutagenesis; mutein; substitution;
XX mutant.

XX Arabidopsis thaliana.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Misc-difference 508

XX /note= "Wild-type Glu substituted by Ala"

XX W0200028053-A2.

XX 18-MAY-2000.

XX 04-NOV-1999; 99WO-US25950.

XX 05-NOV-1998; 98US-0107275.

XX

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX

XX Allen SM, Huang LL, Falco SC, Rafaleki AJ;

XX WPI; 2000-376564/32.

XX Novel polynucleotides encoding plant glutamine amidotransferase

PT homologues, useful for producing transgenic plants and as probes or

PT primers

XX Example 9; Page -: 53pp; English.

XX The present sequence is a mutant of the mouse ear cress HisHf enzyme,

CC which is involved in histidine biosynthesis. It was produced by

CC site-directed mutagenesis, which was carried out to find the active

CC sites within the enzyme, which is made up of a cyclase and a glutamine

CC amidotransferase. The mutation in this case is in the cyclase

CC domain. Determining the active sites is useful for the identification of

CC herbicides and fungicides which inhibit the enzyme. This residue was

CC shown to form part of the active site.

CC Note: The present sequence is not shown in the specification but is

CC derived from the Arabidopsis thaliana wild-type His HF sequence given in

CC the sequence listing (see AAY94231).

XX

SQ Sequence 593 AA;

Query Match 74.9%; Score 2154; DB 21; Length 593;

Best Local Similarity 76.5%; Pred. No. 8.8e-187;

Matches 413; Conservative 60; Mismatches 63; Indels 4; Gaps 3;

QY 21 AGAGDSVVTLLDYGAGNVRNNAIRTLGFDIKDVQKPEDILNAKRLIFPGVGAFAPAM 80

DB 56 ASSTSDSVTLLDYGAGNVRNNAIRTLGFDIKDVQKPEDILNAKRLIFPGVGAFAPAM 115

QY 81 DVLRKGLAELCTYQNDPRPFLGICGLQLLPESSENGPIQGLIPGRVRFSSNG 140

DB 116 DVLRKGLAELCTYQNDPRPFLGICGLQLLPESSENGPIQGLIPGRVRFSSNG 175


```

176 IRVPHIGNALQVGDSEILDDVGNRHVYFVHSYRAIPSDENKDWISSTCNYSGESISSI 235
200 QKGNVHAVQPHPEKSGVGLSILRRFLNADSFNNKQKPMNGKASKAKRVIACLDVRAN 259
236 RRGNVHAVQPHPEKSGVGLSVLRFLHPKL--PATQKPMGKASKAKRVIACLDVRTN 293
260 DNGDLVVTGDDQYDVRERTEENEVRNLGKPVLAGQVYLDGADSVFLNITGFRDPLGD 319
294 DKGDLVVTGDDQYDVRERTEENEVRNLGKPVLAGQVYLDGADSVFLNITGFRDPLGD 353
320 LPMQLVLRASENVFVPLTVGGGIRFTDANGRIYSSLEVASFYRSGADKVSIGSDAVY 379
354 LPMQLVLRQTSKNVFLTVGGGIRFTDANGRIYSSLEVASFYRSGADKVSIGSDAVF 413
380 TABEYIKTGKTKSSIEQISTVYGNQAVVSTDPRRVYLRKDEVEFKAIVKSHHPGNG 439
414 AAEFIKSGVTKGKSSLEQISRVYGNQAVVSTDPRRVYLRKDEVEFKAIVKSHHPGNG 473
440 EEWYQCTVNGGREGPIGAYELAKAVEELGAGEILLNCIDCGGKGFDDILKILSD 499
474 EEWYQCTVNGGREGPIGAYELAKAVEELGAGEILLNCIDCGGKGFDDILKILSD 533
500 AVNIPVIASSGAGVADHFSFVFNETNASALAAGIF-HRKEVPIKAVKEHLKKEGIEVRL 558
534 SVGIPVIASSGAGTPDHFSEVFEDKRICRACCRHPPPERGYQSQVKEHLQERIEVRI 593

RESULT 14
AA94243
ID AAY94243 standard; Protein; 593 AA.
AC AAY94243;
XX
XX
08-AUG-2000 (first entry)
DT
DE Cress mutant cyclase, D447A substitution.
KW Glutamine amidotransferase; histidine biosynthesis; cyclase;
KW mouse ear cress; site-directed mutagenesis; mutein; substitution;
KW mutant.
XX
XX Arabidopsis thaliana.
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 447
FT /note= "Wild-type Asp substituted by Ala"
XX
XX WO200028053-A2.
XX
XX 18-MAY-2000.
XX
XX 04-NOV-1999; 99WO-US25950.
XX
XX 05-NOV-1998; 98US-0107275.
XX
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX Allen SM, Huang LL, Falco SC, Rafaleki AJ;
XX
XX WPI; 2000-376564/32.
XX
XX Novel polynucleotides encoding plant glutamine amidotransferase
XX homologues, useful for producing transgenic plants and as probes or
XX primers.
XX
XX Example 9; Page -: 53pp; English.
XX
XX The present sequence is a mutant of the mouse ear cress HisHIF enzyme,
XX which is involved in histidine biosynthesis. It was produced by
XX site-directed mutagenesis, which was carried out to find the active
XX sites within the enzyme, which is made up of a cyclase and a glutamine
```

```

CC amidotransferase. The mutation in this case is in the cyclase
CC domain. Determining the active sites is useful for the identification of
CC herbicides and fungicides which inhibit the enzyme. This residue was
CC shown to form part of the active site.
CC Note: The present sequence is not shown in the specification but is
CC derived from the Arabidopsis thaliana wild-type His HF sequence given in
CC the sequence listing (see AAY94231).
XX
XX Sequence 593 AA;
SQ
Query Match 74.8%; Score 2152; DB 21; Length 593;
Best Local Similarity 76.5%; Pred. No. 1.3e-186;
Matches 413; Conservative 60; Mismatches 63; Indels 4; Gaps 3;
QY 21 AGAGDGSVVTLLDYGAGNVRSVRNARTIGFDIKDQKPEDILNAKRLIFPGVGAFAPAM 80
DB 56 ASTSDSVVTLLDYGAGNVRSVRNALRHGFSIKDQVTFGDILNADRLIFPGVGAFAPAM 115
QY 81 DVLIRKGLAEALCTYIQNDRPFLGICLGLQLLFESSENGPIQGLGLIFCRVGRFPSSNG 140
DB 116 DVLNRTGMAELCKYIENDRPFLGICLGLQLLFDSSEQNGPVKGLGVIPIGVGRFPDASAG 175
QY 141 LRVPHIGHALDIKESGAILDDVGNRHVYFVHSYRA-NAEDNKKEWISSTCNYSGESISSI 199
DB 176 IRVPHIGNALQVGDSEILDDVGNRHVYFVHSYRAIPSDENKDWISSTCNYSGESISSI 235
QY 200 QKGNVHAVQPHPEKSGVGLSILRRFLNADSFNNKQKPMNGKASKAKRVIACLDVRAN 259
DB 236 RRGNVHAVQPHPEKSGVGLSVLRFLHPKL--PATQKPMGKASKAKRVIACLDVRTN 293
QY 260 DNGDLVVTGDDQYDVRERTEENEVRNLGKPVLAGQVYLDGADSVFLNITGFRDPLGD 319
DB 294 DKGDLVVTGDDQYDVRERTEENEVRNLGKPVLAGQVYLDGADSVFLNITGFRDPLGD 353
QY 320 LPMQLVLRASENVFVPLTVGGGIRFTDANGRIYSSLEVASFYRSGADKVSIGSDAVY 379
DB 354 LPMQLVLRQTSKNVFLTVGGGIRFTDANGRIYSSLEVASFYRSGADKVSIGSDAVF 413
QY 380 TABEYIKTGKTKSSIEQISTVYGNQAVVSTDPRRVYLRKDEVEFKAIVKSHHPGNG 439
DB 414 AAEFIKSGVTKGKSSLEQISRVYGNQAVVSTDPRRVYLRKDEVEFKAIVKSHHPGNG 473
QY 440 EEWYQCTVNGGREGPIGAYELAKAVEELGAGEILLNCIDCGGKGFDDILKILSD 499
DB 474 EEWYQCTVNGGREGPIGAYELAKAVEELGAGEILLNCIDCGGKGFDDILKILSD 533
QY 500 AVNIPVIASSGAGVADHFSFVFNETNASALAAGIF-HRKEVPIKAVKEHLKKEGIEVRL 558
DB 534 SVGIPVIASSGAGTPDHFSEVFEDKRICRACCRHPPPERGYQSQVKEHLQERIEVRI 593

RESULT 15
AA94245
ID AAY94245 standard; Protein; 593 AA.
XX
XX AAY94245;
XX
XX 08-AUG-2000 (first entry)
DT
DE Cress mutant cyclase, D517A substitution.
XX
XX Glutamine amidotransferase; histidine biosynthesis; cyclase;
XX mouse ear cress; site-directed mutagenesis; mutein; substitution;
XX mutant.
XX
XX Arabidopsis thaliana.
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Misc-difference 517
FT /note= "Wild type Asp substituted by Ala"
XX
XX WO200028053-A2.
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:25:12 ; Search time 22 seconds
(without alignments)
1073.157 Million cell updates/sec

Title: US-09-831-233A-2

Perfect score: 2877

Sequence: 1 HEKELASTKQNGFRIRAL.....EVPKAVKEHLLKEGIEVRL 558

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*

2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*

4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*

5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pep.*

6: /cgn2_6/ptodata/1/iaa/backfilese1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	454	15.8	311	US-09-352-991A-23527	Sequence 23527, A
2	442	15.4	268	US-09-328-352-6187	Sequence 6187, Ap
3	430.5	15.0	258	US-09-134-001C-4647	Sequence 4647, Ap
4	365	12.7	207	US-09-328-352-6054	Sequence 6054, Ap
5	357	12.4	192	US-08-953-139-2	Sequence 2, Appli
6	357	12.4	192	US-08-953-139-4	Sequence 4, Appli
7	355	12.3	182	US-09-134-001C-4652	Sequence 4652, Ap
8	351	12.2	371	US-09-352-991A-23484	Sequence 23484, A
9	350	12.2	235	US-09-352-991A-28794	Sequence 28794, A
10	268	9.3	265	US-09-352-991A-28727	Sequence 28727, A
11	264	9.2	251	US-08-846-762-11	Sequence 11, Appl
12	184	6.4	118	US-08-846-762-10	Sequence 10, Appl
13	138	4.8	261	US-09-352-991A-23486	Sequence 23486, A
14	129	4.5	265	US-09-328-352-6115	Sequence 6115, Ap
15	126.5	4.4	243	US-09-134-001C-4654	Sequence 4654, Ap
16	115	4.0	525	US-09-328-352-4897	Sequence 4897, Ap
17	108.5	3.8	881	US-08-960-780-32	Sequence 32, Appl
18	108.5	3.8	881	US-09-073-898-32	Sequence 32, Appl
19	107	3.7	188	US-08-858-207A-445	Sequence 445, App
20	107	3.7	393	US-09-629-616-2	Sequence 2, Appli
21	107	3.7	667	US-08-471-033-7	Sequence 7, Appli
22	107	3.7	667	US-08-471-044-7	Sequence 7, Appli
23	107	3.7	667	US-08-463-483A-7	Sequence 7, Appli
24	107	3.7	667	US-08-471-046A-7	Sequence 7, Appli
25	107	3.7	667	US-08-470-566B-7	Sequence 7, Appli
26	107	3.7	667	US-08-469-334-7	Sequence 7, Appli
27	107	3.7	667	US-09-300-529-7	Sequence 7, Appli

28	107	3.7	852	1	US-08-471-033-36	Sequence 36, Appl
29	107	3.7	852	2	US-08-471-044-36	Sequence 36, Appl
30	107	3.7	852	2	US-08-463-483A-36	Sequence 36, Appl
31	107	3.7	852	2	US-08-471-046A-36	Sequence 36, Appl
32	107	3.7	852	2	US-08-470-566B-36	Sequence 36, Appl
33	107	3.7	852	2	US-08-469-334-36	Sequence 36, Appl
34	107	3.7	852	3	US-09-300-529-36	Sequence 36, Appl
35	107	3.7	884	1	US-08-471-033-5	Sequence 5, Appli
36	107	3.7	884	2	US-08-471-044-5	Sequence 5, Appli
37	107	3.7	884	2	US-08-463-483A-5	Sequence 5, Appli
38	107	3.7	884	2	US-08-471-046A-5	Sequence 5, Appli
39	107	3.7	884	2	US-08-470-566B-5	Sequence 5, Appli
40	107	3.7	884	2	US-08-469-334-5	Sequence 5, Appli
41	107	3.7	884	3	US-09-300-529-5	Sequence 5, Appli
42	107	3.7	1094	4	US-09-268-347-32	Sequence 32, Appl
43	107	3.7	1338	1	US-08-471-033-50	Sequence 50, Appl
44	107	3.7	1338	2	US-08-471-044-50	Sequence 50, Appl
45	107	3.7	1338	2	US-08-463-483A-50	Sequence 50, Appl

ALIGNMENTS

RESULT 1
US-09-252-991A-23527
; Sequence 23527, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23527
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-23527

Query Match	15.8%;	Score 454;	DB 4;	Length 311;
Best Local Similarity	37.5%;	Pred. No. 4.2e-38;	Mismatches 92;	Indels 58; Gaps 8;
Matches 117;	Conservative 45;			
QY	246	LAKRVTACLDVRANDGDLVVTGKDQYDVRETEENEVRNLGKPVLAGQYLLDGADSVS	305	
Db	58	LAKRIIPCLDV--DNGRVV--KGVKFE-----NIRDAGDPVEIARRYDQGADEIT	104	
QY	306	FLNITGRFPDGLPMLQVLRASENVFPLTVGGGIRDFDFTDANGRYYSLEVASEYFR	365	
Db	105	FLDITASVD---GRDFTLTHVERMASQVFIPLTVGGGVSQDIR-----NLLN	150	
QY	366	SGADKVSIGSDAVYTAEEYIKTVKTKSSIEQISTVYGNQAVVVSIDPRVYLRKPDEV	425	
Db	151	AGADKVSINTAAVFNP-----FVGEADRFQSGCIVVAID-----	186	
QY	426	EFKAIKVSHPGNGEYAWYQCTVNGRGGRPTGAYELAKAVEELGAGBILNCIDCDGQ	485	
Db	187	---AKKVSAPG---EAPRWEFTHGKRPKTGLDVLWAKWEDLGAGELLTSMQDGV	239	
QY	486	KGKFDIDLKLSDAVNIPIVASSGAGVADHSEVFENETNSAALAAGTFRKREYPIKAV	545	
Db	240	KSGYDLGVTRAISEAVNVPVIASGGVGNLEHLAAGILEGKADAVLAASIFHFGEYTVPEA	299	
QY	546	KEHLLKEGIEVR	557	
Db	300	KAYLASRGIVVR	311	

RESULT 2
US-09-328-352-6187
; Sequence 6187, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6187
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6187

Query Match 15.4%; Score 442; DB 4; Length 268;
Best Local Similarity 35.0%; Pred. No. 5.7e-37;
Matches 114; Conservative 52; Mismatches 98; Indels 62; Gaps 8;

Qy 233 NKQKPMNGKASKLAKRVIACLDVRANDGDLVVTKGQDYDVRTEENEVRNLGKPV 292
Db 5 NFQRRFLFKLAKLAKLIIICLDV--DGRVV--KGQF-----LDIRAGDPVEV 51

Qy 293 AGQYLDGADVSVFNTITGRDPLPMLQVLRASENVFVPLTVGGGIRDTFANGR 352
Db 52 ARRYNEQCADEITFLDITATHH---GRDITRTVERMAETVFVPLTVGGGVRKVEDIRA- 107

Qy 353 YYSLSVASFYFGADKVGSDAVYTABEYIKTVKTKSSIEQISITVYGNQAVVSI 412
Db 108 -----LLNAGADKVSINSAAVFNPE-----FVQEASQHFQAQCIIVAI 145

Qy 413 DPRRVLYRKDVEVEFKAIVKSHPGNGEYAWYQCTVNGRGREGRPICAYELAKAVBELGA 472
Db 146 DAKT-----GDNK-----WEIFTHGGRKPTGIDAIWAVKMDYGA 182

Qy 473 GEILLNCIDCGQKGFIDILKISDAVNIPVIASSGAGVADHFSEVFNETNASALAA 532
Db 183 GELLITSMADAGTKAGYDIALMRAINDRVTIPTIASSGGVGNLQHLADGILQGGADAVLAA 242

Qy 533 GIFHRKEVPKAVKEHLLKGEIVRL 558
Db 243 SIFHFGQYTIPEAKOYLAEQIEMRL 268

RESULT 3
US-09-134-001C-4647
; Sequence 4647, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4647
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4647

Query Match 15.0%; Score 430.5; DB 4; Length 258;
Best Local Similarity 34.0%; Pred. No. 8.1e-36;
Matches 106; Conservative 53; Mismatches 92; Indels 61; Gaps 7;

Qy 246 LAKRVIACLDVRANDGDLVVTKGQDYDVRTEENEVRNLGKPVLAGQYVLDGADEVS 305
Db 8 IKRVIPLCLDVK-----DGRVVKGFQF-----QSLRDIGNPVDLALYNEAGADELV 54

Qy 306 FLNITGFRDFPLGLPMLQVLRASENVFVPLTVGGGIRDTFANGRYYSLSVEASEYFR 365
Db 55 FLDIS---KTEAGHDLMIIEVATAKQLFIPLTVGGGIGNLDD-----ITQLLN 100

Qy 366 SGADKVSIGSDAVYTABEYIKTVKTKSSIEQISITVYGNQAVVSIIDPRRVLYRKPEV 425
Db 101 HGADKISLNSALKHPE-----LIRQASEKFGRCICIAID----- 136

Qy 426 EFKAIVKSHPGNGEYAWYQCTVNGRGREGRPICAYELAKAVBELGAGBILNCIDCDGQ 485
Db 137 -----SFYDKDREDYF---CTTHGGKKLTDVSVYDWQVEVHGLGAGELLITSMHDDGM 186

Qy 486 GKGFIDILKISDAVNIPVIASSGAGVADHFSEVFNETNASALAAAGIFHRKEVPKAV 545
Db 187 KQGFIDIEHLAKIKQLVNIPIIASGGGNAQHFVELFQOTDVSAGLAASILHDQETTVAEI 246

Qy 546 KEHLLKGEIVR 557
Db 247 KDKMREGGILVR 258

RESULT 4
US-09-328-352-6054
; Sequence 6054, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6054
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6054

Query Match 12.7%; Score 365; DB 4; Length 207;
Best Local Similarity 37.3%; Pred. No. 3.1e-29;
Matches 78; Conservative 35; Mismatches 74; Indels 22; Gaps 5;

Qy 29 VTLLDYGAGNVRSVRNAIRTLGFDIKDVQKPEDILNAKRLIFFGVGAFAPAMDVLIRKGL 88
Db 6 IALLDYGMLHLSAAKALEYVGATVDVTDNDPKLIAQADKIVFPFGVGMRCMQGMREAGI 65

Qy 89 AEALCTYIQRDPFLGICLGLQLLFSSSENGPIQGLGLIPGRVGRFESSNGLRVPHIGW 148
Db 66 DEVVRKAFFN-KPVALICVGMQALLQSSBENGVDALGIPDGIIVKHFPMQMEGLKVPBMGW 124

Qy 149 ----HALD-----IKESGAILDDVGNQHVYVHSHYRANAEDNKEMWISSTCSVGDDFIA 197
Db 125 NQVHQMDPSHPMMWNIEQ-----DARFYVHSYVVEPKD-ENLVAATCEYGVNFT 174

Qy 198 SIOKGNVHAVQFHPKSGGVGLSILRRFL 226
Db 175 AIHKDNLFAQFHPKSHTAGLQLLNKFV 203

RESULT 5
US-08-953-139-2
; Sequence 2, Application US/08953139
; Patent No. 6258578
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin K.
; APPLICANT: Lonetto, Michael A.
; APPLICANT: Warren, Patrick V.

;; TITLE OF INVENTION: NOVEL HIS5
;; NUMBER OF SEQUENCES: 7
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Dechert Price & Rhoads
;; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
;; CITY: Philadelphia
;; STATE: PA

;; COUNTRY: US

;; ZIP: 19103

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette

;; COMPUTER: IBM Compatible

;; OPERATING SYSTEM: DOS

;; SOFTWARE: FastSeq for Windows Version 2.0

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/953,139

;; FILING DATE:

;; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER:

;; FILING DATE:

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Dickinson, Todd Q

;; REGISTRATION NUMBER: 28,354

;; REFERENCE/DOCKET NUMBER: GM10056-01

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 215-994-2252

;; TELEFAX: 215-994-2222

;; TELEX:

;; INFORMATION FOR SEQ ID NO: 2:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 192 amino acids

;; TYPE: amino acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; US-08-953-139-2

Query Match 12.4%; Score 357; DB 3; Length 192;

Best Local Similarity 38.8%; Pred. No. 1.9e-28;

Matches 80; Conservative 35; Mismatches 77; Indels 14; Gaps 5;

QY 28 VTLLDYGAGNVSVRNAITLGFIDKDVQKPEDILNAKELIFPGVGAFAPAMDVLIRKG 87

Db 1 MIVIDVYGLGNISNVKRAIEHLGYEVVVVNTSKIIDQAEITILPFGVGHFADAMSEIKRLN 60

QY 88 LAEALCTYIQNDPFPFICIGLQLFPESSENGPIQGLIPGRVGRFESSNGLRVPHIG 147

Db 61 LNAILAK--NTDKMIGICIGLMQMYEHSDE-GDASGLGFIPGNISRIQTE--YVPVHLG 115

QY 148 WHALDIKESAILDDVGNQHVYFVHSYRANAENKEMWISSTCSYGDDEFIASIQKGNHAV 207

Db 116 WNNLVSKH-----PMLNQDVYFVHSYQAPMSEN---VIAVAYGADIPAIVQFNYYIGI 166

QY 208 QFHPEKSGGVGLSILRRFLNADSFNN 233

Db 167 QFHPEKSGTYGLQLRQAIQGGFIND 192

RESULT 6

US-08-953-139-4

Sequence 4, Application US/08953139

Patent No. 6258578

GENERAL INFORMATION:

APPLICANT: Burnham, Martin K.

APPLICANT: Lonetto, Michael A.

APPLICANT: Warren, Patrick V.

TITLE OF INVENTION: NOVEL HIS5

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dechert Price & Rhoads

STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

CITY: Philadelphia

STATE: PA

;; COUNTRY: US

;; ZIP: 19103

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette

;; COMPUTER: IBM Compatible

;; OPERATING SYSTEM: DOS

;; SOFTWARE: FastSeq for Windows Version 2.0

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/953,139

;; FILING DATE:

;; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER:

;; FILING DATE:

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Dickinson, Todd Q

;; REGISTRATION NUMBER: 28,354

;; REFERENCE/DOCKET NUMBER: GM10056-01

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 215-994-2252

;; TELEFAX: 215-994-2222

;; TELEX:

;; INFORMATION FOR SEQ ID NO: 4:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 192 amino acids

;; TYPE: amino acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; US-08-953-139-4

Query Match 12.4%; Score 357; DB 3; Length 192;

Best Local Similarity 38.8%; Pred. No. 1.9e-28;

Matches 80; Conservative 35; Mismatches 77; Indels 14; Gaps 5;

QY 28 VTLLDYGAGNVSVRNAITLGFIDKDVQKPEDILNAKELIFPGVGAFAPAMDVLIRKG 87

Db 1 MIVIDVYGLGNISNVKRAIEHLGYEVVVVNTSKIIDQAEITILPFGVGHFADAMSEIKRLN 60

QY 88 LAEALCTYIQNDPFPFICIGLQLFPESSENGPIQGLIPGRVGRFESSNGLRVPHIG 147

Db 61 LNAILAK--NTDKMIGICIGLMQMYEHSDE-GDASGLGFIPGNISRIQTE--YVPVHLG 115

QY 148 WHALDIKESAILDDVGNQHVYFVHSYRANAENKEMWISSTCSYGDDEFIASIQKGNHAV 207

Db 116 WNNLVSKH-----PMLNQDVYFVHSYQAPMSEN---VIAVAYGADIPAIVQFNYYIGI 166

QY 208 QFHPEKSGGVGLSILRRFLNADSFNN 233

Db 167 QFHPEKSGTYGLQLRQAIQGGFIND 192

RESULT 7

US-09-134-001C-4652

Sequence 4652, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4652

LENGTH: 192

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4652

Query Match 12.3%; Score 355; DB 4; Length 192; 8;
Best Local Similarity 36.2%; Pred. No. 2.9e-28;
Matches 81; Conservative 39; Mismatches 66; Indels 26; Gaps 8;

QY 28 VVTLLDYGAGNVSVRNAIRITIGFDI-----KDVOKPEDILNAKLIPGVGAFAPAMD 81
DB :
1 MIAIIDVGLNISNVTRAIOHLYGDVILTCCDDKDVK-----AEAIVLPVGWHFQDAMH 54
QY 82 VLIRKGALAEALCTTYQNDRPFLGICLQLPESSENGPIQGLIGLIPGRVGRFESSNGL 141
DB :
55 SIEKSIDMKLKN--HDKEPIIGLCGMQLLFQHSAE-GDVSLELVPGNIVIPQSSH-- 109
QY 142 RVPHIGWHALDIKEGSAILDDVNQHYYFVHSYRANAENKWEISSTCSYGDDFIASIQK 201
DB :
110 PPHLGWN--ELKSTHPLL---QSDVYFVHSYQA---EMSEYWAYADYGTIKPGVIQV 160
QY 202 GNHVAVQFHPEKSGGVGLSILRRFLNADSENN 233
DB :
161 RNVIQGFHPEKSGTGYGLEILNQALKRGFFIND 192

RESULT 8
US-09-252-991A-23484
; Sequence 23484, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23484
; LENGTH: 371
; TYPE: PRP
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23484

Query Match 12.2%; Score 351; DB 4; Length 371;
Best Local Similarity 36.7%; Pred. No. 2.3e-27;
Matches 87; Conservative 38; Mismatches 94; Indels 18; Gaps 6;

QY 9 KPONGFRIRAALAGAGDSV-----VTLDDYGAGNVSVRNAIRTLGFDIKDVQKP 59
DB :
134 RPRAAHGRAGRAGRADGRADAQRVPVMTVAVIDYGMNLHSAKALEHVAGRVLVSSD 193
QY 60 EDIL-NAKLRIIPGVGAFAPAMDVLIRKGLAEALCTTYQNDRPFLGICLQLPFESSEE 118
DB :
194 AAIVREADRVPVFPVGGVGAIRDCAEIRLGF-DALVREVSQLDRPFLGICVMQALLERTSE 252
QY 119 NGPIQGLLIPGRVGRF-----BSSNGLRPVPHIGHALDIKEGSAILDDVGNQ-HVYFVH 172
DB :
253 NDGVDCIGLFPQQVRPFGKDLHEAGEHLKYPHMGWQVSOAVEHPLWHEIPDOARFYFVH 312
QY 173 SYRANAEDNKWEISSSTCSYGDDFIASIQKNHAVQFHPEKSGGVGLSILRRFLNAD 229
DB :
313 SYVIEA-GNPQRVVGHGHYGVDFAAALAEGRSFVQFHPEKSHTHGLQLLQNFVAMD 368

RESULT 9
US-09-252-991A-28794
; Sequence 28794, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136

Db 150 -----WGHVSFSENGTRDMKRSPL-----EWAQALEAGVG 181
Qy 474 EILLNCIDCGOGKFPDILIKLISDAVNIPVLIASSGADVHFSEVFNETNASALAAG 533
Db 182 EIFLNSIDRGVQKGFDPNALVENIASNVHVPVIACGAGSIADLIDLIFERTCVS-AVAG 240
Qy 534 ---IFHRK 538
Db 241 SLFVFHGK 248

RESULT 11
US-08-846-762-11
; Sequence 11, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; APPLICANT: de Kievit, Teresa
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
; TITLE OF INVENTION: of O-Antigen in Pseudomonas Aeruginosa
; FILE REFERENCE: 6580-089
; CURRENT APPLICATION NUMBER: US/08/846,762A
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-08-846-762-11

Query Match 9.2%; Score 264; DB 2; Length 251;
Best Local Similarity 30.8%; Pred. No. 1.1e-18;
Matches 92; Conservative 44; Mismatches 87; Indels 76; Gaps 12;
Qy 248 KRVIACLDVRANDGDLVVTGQDYVRETEENVRNLGKPVLAGQYYLDGAEVSPL 307
Db 4 RRVIPCLLLK--DRG-LVKT-----VKPEKPVGDPIINAIKPEKEVDELILL 50
Qy 308 NITGFRDPLGLDPLMLQVLRASENVFVPLTVGGGIRDFTDANGRYYSLEVASFYFRSG 367
Db 51 DIDASR---LNQEPNYELIAEVAGCEPICYGGIK-----TLEHAETKIFSLG 96
Qy 368 ADKVSIGSDAVYTAEEYIKTKGKSSIEQISTVYGNQAVVVSIDPRVYLKPKDEVEF 427
Db 97 VEKVSINTAALMDL-----SLIRRIADKFGSQSVGSDICKRGF-----135
Qy 428 KAIKVSHPGPNGBEYAVVQCTV---NGREGR--PIGAYELAKAVEELGAGEILLNCIDC 482
Db 136 -----WGHVSFSENGTRDMKRSPL-----EWAQALEAGVGEIFLNSIDR 176
Qy 483 DGOCKGFPDILIKLISDAVNIPVLIASSGADVHFSEVFNETNASALAAG---IFHRK 538
Db 177 DGQKGFDPNALVENIASNVHVPVIACGAGSIADLIDLIFERTCVS-AVAGSLFVFHGK 234

RESULT 12
US-08-846-762-10
; Sequence 10, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; APPLICANT: de Kievit, Teresa
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
; TITLE OF INVENTION: of O-Antigen in Pseudomonas Aeruginosa
; FILE REFERENCE: 6580-089
; CURRENT APPLICATION NUMBER: US/08/846,762A
; CURRENT FILING DATE: 1997-04-30

; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-08-846-762-10

Query Match 6.4%; Score 184; DB 2; Length 118;
Best Local Similarity 36.4%; Pred. No. 5e-11;
Matches 44; Conservative 19; Mismatches 50; Indels 8; Gaps 3;

Qy 107 LGLQLLFESSENGPIQGLGLIPGRVGRFESSNGLRVPHIGHALDIKEGSAILDDVGNQ 166
Db 2 LGLR-----SBEGAEPGLGWDMSVRFERRDRKVPHGMGNQVSPQLEHPILSGINEQ 55
Qy 167 -HVFVHSYRANAEDNKEMISSCSDGDDFIASIQKGNVHAVQFHPKSGGVGLSLRRF 225
Db 56 SRPYFVHSYTMVPKOPDD--ILLSCNYGQKFTAAVARDNVFGFQFHPKSHKFGMLKKNF 114
Qy 226 L 226
Db 115 V 115

RESULT 13
US-09-252-991A-23486
; Sequence 23486, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196, 136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23486
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23486

Query Match 4.8%; Score 138; DB 4; Length 261;
Best Local Similarity 21.4%; Pred. No. 1.1e-05;
Matches 56; Conservative 38; Mismatches 90; Indels 78; Gaps 8;
Qy 252 ACILDVRANDGDLVVTGQDYVRETEENVRNLGKPVLAGQYYLDGAEVSFLNITG 311
Db 29 ACVRLRQGLMEDATVFSDD-----PVMAAKWVDGGRRLHLDVLDNG 70
Qy 312 -FRDPLGLDPLMLQVLRASENVFVPLTVGGGIRDFTDANGRYYSLEVASFYFRSGADK 370
Db 71 APEGKVPNGEVVTAIARRYPD---LPIQGGIR-----SLETHVYVRAGVSY 116
Qy 371 VSIGSDAVYTAEEYIKTKGKSSIEQISTVYGNQAVVVSIDPRVYLKPKDEVEFK 428
Db 117 VIIGTKAV-KQPEFVGEACRAPPK-----VIVGLDAKQGV-----152
Qy 429 AIKVSHPGPNGBEYAVVQCTVNGREGRPIGAYELAKAVEELGAGEILLNCIDCGGKG 488
Db 153 -----ATDGAESVSEVQVIDLARRFEADGVSIAVYTDISKGMQMG 193
Qy 489 FOIDLKILSDAVNIPVIASSG 510
Db 194 CNVQATAALANATRIPIVASSG 215

RESULT 14

Db 64 XKLIPGVGHGHCLSQASAGYMAPIRAHIESGKPFMGICVGLQALFEGSSDPNCPGL 123
Qy 126 GLIPGRVGRSESSNGLRVPHIGWHALDIKESAILDDVGNQHYVYVHSYR-ANAEDN 181
Db 124 GVLPGKLDREDDTS-KSVPHIGMNDASCPNPNLGLNPSKYVYVHSYKMPYTKGOLES 182
Qy 182 KEWISSTCSY-GDDFIASIOKGNHVAQVPHPEKSGVGLSLRRFLN-----ADSFNNKR 235
Db 183 QGNAVATVYGGTEFFIGAIAGNVMATQPHPEKSGVAGLVRIRAFLDGSGAALASHPPQ 242
Qy 236 QKPMNGK---ASK--LAKRVIACLDVRANDNGDLVVTGQDYDVRETEENEVRNLKPV 290
Db 243 EVPNAGDVALYSKEGLTRRVIACLDVRTNDQGLVVTGQDYDVREKSDDRNVRNLKPV 302
Qy 291 ELAQVYLDGADSVFNITGPRDPLGLPMLQVLRASENVFVPLTVGGGTRDFTDAN 350
Db 303 EMARKYEQGADEVFTFNITSFRCPPVADLPMLBILRLTSTKVTVPLTVGGGTRDFTD 362
Qy 351 GRYVSSLEVASSEYFRSGADKVSIGSDAVYTAEEYIKTGVT--GKSSIEQISTVYGNQAV 408
Db 363 GTKVSALEIATMYFQSGADKVSIGSDAVIAEEYASG-KTLFGNTAIEQISKAYGNQAV 421
Qy 409 VVSDPRVYLRKPDEVEPKAKVSHPGPNGEYAMVYQCTVNGGREGRPIGAYELAKAVE 468
Db 422 VVSDPRVYLRKPDEVEPKAKVSHPGPNGEYAMVYQCTVNGGREGRPIGAYELAKAVE 481
Qy 469 ELGAGEILLNCIDCGOGKGFDDILIKLSDAVNIPVIASSGAGVADHSEVFNETNASE 528
Db 482 AMGCGEILLNCIDKDGNSGFDLEIRQVKAARIPVIASSGAGNPGHFEVFRFTTDA 541
Qy 529 ALAAGIFHRKEVPIKAVEHLLKEGIEVR 557
Db 542 ALGAGMFRGEYTVQVQKEELKARGLVVR 570

RESULT 2
US-10-369-493-1500
; Sequence 1500, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1500
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1500

Query Match 47.4%; Score 1363; DB 12; Length 552;
Best Local Similarity 50.4%; Pred. No. 2,9e-123;
Matches 281; Conservative 91; Mismatches 149; Indels 36; Gaps 11;
Qy 28 VVTLDDYGAGNVRVRNAITFLGPDIDKQKPED--ILNAKRLIFPGVGAFAPAMVILIR 85
Db 3 VVHVIVDESGLQSLTNAIEHLGVEVLVKSQKDFNISGTSRLTLPVGNYGHFVDNLFN 62
Qy 86 KGLAEALCTYQNDPRFLICGLQLLPESSENGPIQGLIPGRVGRSESSNGLRVPH 145
Db 63 RGEFKPIREYESGKPTMGICVGLQALFAGSVESPKSTGLNIDFKLSRFDDBS-KVPE 121
Qy 146 IGMHALDIKESGAI-LDDVGNQHYVYVHSYRA--NAE-----DNKEWISSTCSYG-DDFI 196

Db 122 IGNNSCIPSENLPFGLDPY--KRYVYFVHGFAAILNSEKKKNLENDGWKIAKAKYGSBEFI 179
Qy 197 ASIOKGNHVAQVPHPEKSGVGLSLRRFLNADSFNNKROK-----MNGKA 243
Db 180 AAVNKNITFATQPHPEKSGKAGLNVTENFL-----KQSPPIPNYSABEELLMDYS 232
Qy 244 S-KLAKRVIACLDVRANDNGDLVVTGQDYDVRETEENEVRNLKRPVELAGQYVYLDGAD 302
Db 233 NYGLTRRIIACLDVRTNDQGLVVTGQDYDVREKSDGKGRNLGKPVQLAQKYQQGAD 292
Qy 303 EYVFLNITGPRDPLGLPMLQVLRASENVFVPLTVGGGTRDFTDANGRYVSSLEVASSE 362
Db 293 EVTFNLITSPRDCPLKDTMPLVLEKQAARTVFPVPLTVGGGIDKIDVDVDTGKIPALEVASL 352
Qy 363 YFRSGADKVSIGSDAVYTAEEYIKTGVT--GKSSIEQISTVYGNQAVVVSIDPRRVYLRK 421
Db 353 YFRSGADKVSIGSDAVYTAEEYIKTGVT--GKSSIEQISTVYGNQAVVVSIDPRRVYLRK 412
Qy 422 PDEVEFKAKVSHPGPNGEYAMVYQCTVNGGREGRPIGAYELAKAVEELGAGELLNCID 481
Db 413 QADTKNKVETETEPGNGEKYCWYQCTIKGGRSRLDGVWELTRACEALGAGELLNCID 472
Qy 482 CDGQGGKGFDDILIKLSDAVNIPVIASSGAGVADHSEVFNETNASEAALAGIFHRKEVP 541
Db 473 KDGNSGYDLELIEHVKDAVKIPVIASSGAGVPEHFEAFKTRADACLGAGMFRHGEFT 532
Qy 542 IKAVEHLLKEGIEVRL 558
Db 533 VNDVKBYLLEHGLKVRM 549

RESULT 3
US-10-369-493-22552
; Sequence 22552, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22552
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-22552

Query Match 46.1%; Score 1327.5; DB 12; Length 502;
Best Local Similarity 52.4%; Pred. No. 6.9e-120;
Matches 268; Conservative 82; Mismatches 128; Indels 33; Gaps 8;
Qy 68 LIFPGVGAFAPAMDVILIRKGLAEALCTYQNDPRFLICGLQLLPESSENGPIQGLGL 127
Db 3 LIFPGVGNFGVCDLSLAKQGLFELRRYALSGKPFMAVCVIGIQLFEGSVAPHSKGLGV 62
Qy 128 IPRVGRFESSNGLRVPHIGWHALDIKESAILDDVGNQ-----HVFVHSYRANAE 179
Db 63 FPLGVQRFNDND-KTVPHIGWNS-----CAVRSDTSKFEFGMRPHDKPFVHVSYMIPEK 115
Qy 180 D---NKEWISSTCSYGDD-PIASIOKGNHVAQVPHPEKSGVGLSLRRFLNADSFNNKR 235
Db 116 GLILPPEFKIATTKYGNFVGAIVKQNFATQPHPEKSGSAGRLCKAFLTG-----Y 170
Qy 236 QKPMNGKASK-----LAKRVIACLDVRANDNGDLVVTGQDYDVRETEENEVRNLG 287

; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1220
; LENGTH: 301
; TYPE: PRP
; ORGANISM: Methanobacterium thermoautotrophicum
US-10-369-493-1220

Query Match 18.1%; Score 522; DB 12; Length 301;
Best Local Similarity 40.6%; Pred. No. 5.8e-42;
Matches 127; Conservative 44; Mismatches 102; Indels 40; Gaps 7;

QY 246 LAKRVIAACLDVRANDGDLVVTGDDQYDVRERTEENVRNLGKPVLAGQYVLDGAEV 305
DB 29 LAKRIIPCLDC-----DLQVPG--RVVKGVEFKQIRVAGDPVELATRYEDGAEIV 79

QY 306 FLAITGRDFPLGDLPLMLQVLRASENVFVPLTVGGGIRDFTDANGRYVSSLEVAS 365
DB 80 FLDITASHE---RRETWTHVIEATTENVFVPCVGGGIRKPED-----YFKWL-----K 125

QY 366 SGADKVSIGDVAVTAEYIKTGKTSIEQISTVYGNQAVVVSIDPRRVYLKRPDEV 425
DB 126 AGADKCSNTAAKNPE-----LINEASDLVGSQACVVAIDAKERYIENPRES 173

QY 426 EFKAIVSHPGPNGEYAWYQCTVNGREGRPIGAYELAKAVBELGAGEILLNCIDCDGQ 485
DB 174 DERPIEVDG-----YCWYECSTYGGREFTGIDAVKAMECDQDQAGEILLTSMRDGT 228

QY 486 GKGFDDILKISDANVPIVSIASGAGVADHSEVFNETHASALAAGIPIHKEVPIKAV 545
DB 229 KMGYDIFLRTMSNLDPIVSIASGAGVEPIHVEAFTDGDADAALAASIFHFNEYVPVAV 288

QY 546 KEHLLEKIEVRL 558
DB 289 KEYLRSRGVIRL 301

RESULT 7
US-10-369-493-20868
; Sequence 20868, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20868
; LENGTH: 438
; TYPE: PRP
; ORGANISM: SYNECHOCOCCUS SP. WH 8102
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(438)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20868

Query Match 17.7%; Score 510.5; DB 12; Length 438;
Best Local Similarity 29.9%; Pred. No. 1.4e-40;
Matches 161; Conservative 79; Mismatches 182; Indels 117; Gaps 19;

QY 27 SVVTLDDYAGNVRSVRNAIRTLGFDIKDQVKPDIINAKRLIFPGVGAFAPAMDVLIRK 86

DB 1 SKIAIVDYGMCNLWSIKSAIQIGVDSILTSDDPKDILNSSAILPGVGSFKTKMDNLLSL 60
QY 87 GLAEL---CTYIQNDRRPFLGICLGLIFSESENGPIQGLIGLIPRV---GRESSN 139
DB 61 GLSQAIIDACMP--RSIPILGICLGFQLCCSEEPSYTKGLSLPLPIQIVPLCRHIDAS- 117
QY 140 GLRVPHIGHALDIKESAILDDVGNQ--HVYFVHSYRANAEDNKWISSTCSYGGDDFIAS 198
DB 118 -FVLPVHGFTSVTTSKNDPLFNIAKNSDFYFVHSYGAFNVPHDFTTYSCYNDAKIIS 176
QY 199 IOKGNVHAVQFHPKSGGVGLSILRFLNADSFNNKQKPMNGKASKLAKRVIAACLDVRA 258
DB 177 ANVNHIMGVQFHPKSGKQSTNGLLLDNLF---SFSNV-----XKAYYLV--LLYC----- 220
QY 259 NNGDLVVTK-----GQYDVRERTEENVRNLGKPVLAGQYVLDGAEVSLNITG 311
DB 221 --DGFFCLSRNFKLQIGDFRMLQRYNFN-----YSATTFIDELIILDISR 264
QY 312 FRDFPLGDLPLMLQVLRASENVFVPLTVGGGIRDFTDANGRYVSSLEVASSEYFRSGADKV 371
DB 265 -KSRDINKLPPVYL---SENFVPIVITAGGIRSFEF-----AKVLENGADKV 309
QY 372 SIGSDAVYTAEEYIKTGKTSIEQISTVYGNQAVVVSIDPRRVYLKRPDEVPEPKAIK 431
DB 310 CLNTSLIQCPH-----VSEKISSVYQQQSLVASID-----FKHDHGDPKFF- 350
QY 432 VSHPGNGSEYAWYQCTVNGREGRPIGAYELAKAVBELGAGEILLNCIDCDGQKGFDI 491
DB 351 -----IDNGLIEVOYNIQELISFLDPLPFCCILLOSVDRTDGTGTGFDL 393
QY 492 DLIKLISDANVPIVSIASGAGVADHSEVFNETHASALAAGIPIHKEVPIKAVKEHLL 550
DB 394 TLANTPRDQSLRPILLGGAGHSDH-----LVEGLLHGSTDAVATA--HLL 437

RESULT 8
US-10-369-493-21460
; Sequence 21460, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21460
; LENGTH: 272
; TYPE: PRP
; ORGANISM: Methanococcus jannaschii
US-10-369-493-21460

Query Match 17.4%; Score 499.5; DB 12; Length 272;
Best Local Similarity 40.1%; Pred. No. 7.5e-40;
Matches 128; Conservative 47; Mismatches 89; Indels 55; Gaps 10;

QY 246 LAKRVIAACLDVRANDGDLVVTGDDQYDVRERTEENVRNLGKPVLAGQYVLDGAEV 304
DB 2 LTKRIIPCLDIK-----DGRVVKG-----TKFLNLRDAGDPVELA-QYDDEGADEL 47

QY 305 SFNLITGF---RDFPLGDLPLMLQVLRASENVFVPLTVGGGIRDFTDANGRYVSSLEVAS 361
DB 48 VFLDITASAEKRII-----IIDVVERTAEKVPIPLTVGGGIK-----SIEDFR 90

QY 362 EYFRSGADKVSIGSDAVYTAEEYIKTGKTSIEQISTVYGNQAVVVSIDPRRVYLK 421

Db 91 RILRAGADKVSINTAAV-----KNPLIKEASEIFGSCVVAIDAKRHYNE 138
QY 422 P--DEVEFRAIKVSHPGNGEEVAYQCTVNGREGRPIGAYELAKAVEELGAGEILLNC 479
Db 139 DEIDKINKNVKVE-----DVCWFVYIGRKETGIDAINWAKKVELGAGEILLTS 192
QY 480 IDCQGGKGFIDILKILSDAVNIPVIASSGAGVADHFSEVFNETNASAAAGIFHRKE 539
Db 193 IDKGTGKSGVDLITLTKREISKVKLPVIASGGCGKPEHVYEFYVGKADAALMAGILHYRE 252
QY 540 VPIKAVKEHLKKEGIEVRL 558
Db 253 YTIEEIKKYCADRGIPML 271

RESULT 9

US-10-369-493-11321
; Sequence 11321, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11321
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Methanosarcina mazei
US-10-369-493-11321

Query Match 16.9%; Score 486.5; DB 12; Length 273;
Best Local Similarity 37.8%; Pred. No. 1.4e-38;
Matches 119; Conservative 51; Mismatches 100; Indels 45; Gaps 7;
QY 246 LAKRVIACLDVRANDGDLVVTGQDYDVRETEENEVRNLGKPVLAGQYVLDGADSVS 305
Db 2 LTRKIIPCLDVTDRAAGCVKGVFVDLKE-----AGDPVELAKRYNEEGADELV 52
QY 306 FLNITGFRDPLGLDPLQLQVLRASENVFVPLTVGGGIRDFTDANGRYYSLSLEVASEYFR 365
Db 53 FLDTITASAQ---GRKTMIDVIERTADEVFPLTIGGI-----NSIDAIRQILR 98
QY 366 SGADKVSIGSDAVYTAEBYIKTGVTGKSSIEQISTVYGNQAVVSDIPRR--VYLKPD 423
Db 99 AGADKVSVNTSAVKNP-----FIKSSDIFGAQCIIVTADICRRNTDIKNPD 146
QY 424 EVEFKAIVSHPGNGEEVAYQCTVNGREGRPIGAYELAKAVEELGAGEILLNCIDCD 483
Db 147 KT---ILEEDGTG-----AWYEVYIGGREATGIDAVQWARKAEELGSGEILLTSMRD 198
QY 484 GQGGKGFIDILKILSDAVNIPVIASSGAGVADHFSEVFNETNASAAAGIFHRKEVPK 543
Db 199 GTCAGYDLPITKLSBELDIPVIASGGVGNPQHIYEGFSIGKADAALAAASIFHREYSIK 258
QY 544 AVKEHLKKEGIEVRL 558
Db 259 EVKEYLRDRREIPVRL 273

RESULT 10

US-09-738-626-5791
; Sequence 5791, Application US/09738626
; Publication No. US20020197605A1

; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5791
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-5791

Query Match 16.7%; Score 481.5; DB 10; Length 258;
Best Local Similarity 39.4%; Pred. No. 3.9e-38;
Matches 123; Conservative 49; Mismatches 83; Indels 57; Gaps 10;
QY 246 LAKRVIACLDVRANDGDLVVTGQDYDVRETEENEVRNLGKPVLAGQYVLDGADSVS 305
Db 3 VAIRVIPCLDV---DNGRVV---KGVNFE-----NLKADGDPVELAKRYNEEGADELT 49
QY 306 FLNITGFRDPLGLDPLQLQVLRASENVFVPLTVGGGIRDFTDANGRYYSLSLEVASEYFR 365
Db 50 FLDTITASKH---GRGTMLDVVRRTADQVFIPLTVGGVR-----SEEDVDQLLR 95
QY 366 SGADKVSIGSDAVYTAEBYIKTGVTGKSSIEQISTVYGNQAVVSDIPRRVYLKRPDEV 425
Db 96 AGADKVSVNTSAIARPE-----LLSELSKRFCAQCIIVLSVDARRV---PEG- 138
QY 426 EFKAIKVSHPGNGEEVAYQCTVNGREGRPIGAYELAKAVEELGAGEILLNCIDCDQ 485
Db 139 -----GTPQPSG-----FEVTHGGSKSAELDAIEWAKRGEELGVGEILLNSMDGDT 186
QY 486 GKGGFIDILKILSDAVNIPVIASSGAGVADHFSEVFNETNASAAAGIFHRKEVPKAV 545
Db 187 KNGFDLELEKVRAAVSIPIVIASGGAGKAEHPFPAV-AAAGANAVLAATIFHFREVTIAEV 245
QY 546 KEHLKKEGIEVR 557
Db 246 KGAIKDAGFEVR 257

RESULT 11

US-10-369-493-18588
; Sequence 18588, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28

;; PRIOR APPLICATION NUMBER: US 60/360,039
;; PRIOR FILING DATE: 2002-02-21
;; NUMBER OF SEQ ID NOS: 47374
;; SEQ ID NO 18588
;; LENGTH: 273
;; TYPE: PRT
;; ORGANISM: Halobacterium sp. NRC-1
US-10-369-493-18588

Query Match 16.6%; Score 478.5; DB 12; Length 273;
Best Local Similarity 36.4%; Pred. No. 8.3e-38;
Matches 116; Conservative 45; Mismatches 99; Indels 59; Gaps 8;

QY 246 LAKRVIACLDVRANDGDLVVTGGDYDVRTEENRNLGKPVLAGOYLLDGADEVS 305
DB 3 LTRKVPICIDVDLDGGEPAVYGVNFE-----ELATGDPVEMAKRYNAGADEFV 54
QY 306 FLNITGPRDFPLGLDPLMLQVLRASENVFPLTVGGGIRDFTDANGRYYSLSLEVASEYFR 365
DB 55 FLDITASAE---GRETMLDTVSAVADEVFPLTVGGGIRTDDIR-----ETLR 100
QY 366 SGADKVSIGSDAVYTAEEYIKTGVTGKSSIEQISTVYGNQAVVVSIDPRRVYLRKPDEV 425
DB 101 AGADKVSINGAI-----ADPSLVDRGAKAFSGQSCIVISVDARRF---DE- 143
QY 426 EFKAIVKSHPGNGEY-----AWYQCTVNGREGRPIGAYELAKAVEELGAGETLLN 478
DB 144 -----QGHVTVQDVGSCFECTVHGREGTGMDALEWQEAQRGAGELFVN 191
QY 479 CIDCGGKGFDDILKILSDAVNIPVIASSGAGVADHSEFNEVNETNASALAAGIFHRK 538
DB 192 SIDADGTQDGVYPLTAAVCDVSTPVIASSGCGAGDMADAV-DAGADAALAASIFHD 250
QY 539 EVPIKAVKEHLKEGIEVR 557
DB 251 EYSIAETKETLADAGYPIR 269

RESULT 12
US-10-369-493-8203
;; Sequence 8203, Application US/10369493
;; Publication No. US20030233675A1
;; GENERAL INFORMATION:
;; APPLICANT: Cao, Yongwei
;; APPLICANT: Hinkle, Gregory J.
;; APPLICANT: Slater, Steven C.
;; APPLICANT: Goldman, Barry S.
;; APPLICANT: Chen, Xianfeng
;; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;; FILE REFERENCE: 38-10(52052)B
;; CURRENT APPLICATION NUMBER: US/10/369,493
;; PRIOR FILING DATE: 2003-02-28
;; PRIOR APPLICATION NUMBER: US 60/360,039
;; PRIOR FILING DATE: 2002-02-21
;; NUMBER OF SEQ ID NOS: 47374
;; SEQ ID NO 8203
;; LENGTH: 256
;; TYPE: PRT
;; ORGANISM: Thermobifida fusca
US-10-369-493-8203

Query Match 16.4%; Score 472; DB 12; Length 256;
Best Local Similarity 39.1%; Pred. No. 3.2e-37;
Matches 122; Conservative 43; Mismatches 89; Indels 58; Gaps 10;

QY 246 LAKRVIACLDVRANDGDLVVTGGDYDVRTEENRNLGKPVLAGOYLLDGADEVS 305
DB 3 LARVIFCLDV---DGRVV--KGVNF-----QNLRDAGDPVELARYDAGADELT 49
QY 306 FLNITGPRDFPLGLDPLMLQVLRASENVFPLTVGGGIRDFTDANGRYYSLSLEVASEYFR 365
DB 50 FLDVTASS---NRETTYDVRRTAEQVFIPLTVGGGV-----STEDVDRLLR 95

QY 366 SGADKVSIGSDAVYTAEEYIKTGVTGKSSIEQISTVYGNQAVVVSIDPRRVYLRKPDEV 425
DB 96 AGADKVSINTAAVRPE-----LIREIAQFGSGLVLSADVRRT-----V 136
QY 426 EFKAIVKSHPGNGEYAWYQCTVNGREGRPIGAYELAKAVEELGAGETLLNCIDCGQ 485
DB 137 NGTAT-----PSG-----FEITTHGGROGTGIDAVEMVQQABELGAGEILLNSMDADGT 185
QY 486 GKGFDDILKILSDAVNIPVIASSGAGVADHSEFNEVNETNASALAAGIFHRKEVPIKAV 545
DB 186 KSGFDLELIRAVRKAVNVPVIASSGAGAVEHPAPAV-DAGANAVLAASVFHGEITISDV 244
QY 546 KEHLKEGIEVR 557
DB 245 KAELEKAGYVPR 256

RESULT 13
US-10-369-493-13786
;; Sequence 13786, Application US/10369493
;; Publication No. US20030233675A1
;; GENERAL INFORMATION:
;; APPLICANT: Cao, Yongwei
;; APPLICANT: Hinkle, Gregory J.
;; APPLICANT: Slater, Steven C.
;; APPLICANT: Goldman, Barry S.
;; APPLICANT: Chen, Xianfeng
;; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;; FILE REFERENCE: 38-10(52052)B
;; CURRENT APPLICATION NUMBER: US/10/369,493
;; CURRENT FILING DATE: 2003-02-28
;; PRIOR APPLICATION NUMBER: US 60/360,039
;; PRIOR FILING DATE: 2002-02-21
;; NUMBER OF SEQ ID NOS: 47374
;; SEQ ID NO 13786
;; LENGTH: 256
;; TYPE: PRT
;; ORGANISM: Pseudomonas fluorescens
US-10-369-493-13786

Query Match 16.2%; Score 465; DB 12; Length 256;
Best Local Similarity 38.1%; Pred. No. 1.5e-36;
Matches 119; Conservative 46; Mismatches 89; Indels 58; Gaps 8;

QY 246 LAKRVIACLDVRANDGDLVVTGGDYDVRTEENRNLGKPVLAGOYLLDGADEVS 305
DB 3 LAKRIIPCLDV---DNGRVV--KGVKFE-----NIRDAGDPVEIARRYDEQGADEIT 49
QY 306 FLNITGPRDFPLGLDPLMLQVLRASENVFPLTVGGGIRDFTDANGRYYSLSLEVASEYFR 365
DB 50 FLDITASVD---GRDTTLHTVVERMASQVFIPLTVGGGVTVQDIR-----NLLN 95
QY 366 SGADKVSIGSDAVYTAEEYIKTGVTGKSSIEQISTVYGNQAVVVSIDPRRVYLRKPDEV 425
DB 96 AGADKVSINTAAVRPE-----FVGEAAQFGSQCIIVAI----- 131
QY 426 EFKAIVKSHPGNGEYAWYQCTVNGREGRPIGAYELAKAVEELGAGETLLNCIDCGQ 485
DB 132 ---AKKVSFGF---ETPRWEIPTHGGRKPTGLDAVEWAKQMEGLGAGEILLTSMQDGM 184
QY 486 GKGFDDILKILSDAVNIPVIASSGAGVADHSEFNEVNETNASALAAGIFHRKEVPIKAV 545
DB 185 KNGFDLGVTRALSALGIPVIASSGGVGNLQHLADGILEGHASAVLAASFIHGEITYVQEA 244
QY 546 KEHLKEGIEVR 557
DB 245 KAYMSKRGIVMR 256

RESULT 14
US-10-369-493-16690

```

; Sequence 16690, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barty S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICRO
; TITLE OF INVENTION: PLANTS WITH IMPRO
; FILE REFERENCE: 38-1052052/B
; CURRENT APPLICATION NUMBER: US/10369,
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,03
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 16690
; LENGTH: 251
; TYPE: PRT
; ORGANISM: Bacillus thuringiensis
US-10-369-493-16690

```

Query Match	15.9%	Score	457;	DB	12;	Length	251;
Best Local Similarity	38.1%;	Pred.	No. 8.9e-36;				
Matches	119;	Conservative	41;	Mismatches	90;	Indels	62; Gaps 7;
QY	246	LAKRVTIACLDVRANDNGDLVVTGKDGYDVRTEENEVRNLGKPVLAGOYYILDGADEVS	305	DB	2	LAKRIIPLCLDVKEGR-----VKGVNPF-----IGLQDVGNPVEIAEAYNEAGADEIV	48
		: :				: : :	
QY	306	FLNITGRFPPLGLDPLMLQVLORASENVFPLTVGGGIRDFTDANGRYYSLEVASSEYFR	365	DB	49	FLDITATHE---GRKTIIDVVEKTAARKVFPLTVGGGI-----SSVKDMYNLLR	94
		: :				: :	
QY	366	SGADKVSYISGDVAITYTAAEEYIKTGVTGKTGKSIEIQISTVYGNOAVVUSDPRRVYLKRPEVD	425	DB	95	AGADKVVINSAAV-----RNPKLTIEGAEHFGSQCIWVAIDARKV-----	134
		: :				: : :	
QY	426	EFKAIKVSHPGPNGEERYAWQCCTVNGREGREPTICAVELAKAVELGAGEILLNCIDCDQG	485	DB	135	-----AEGKW-NVTYNGRGVRTGMDAIEWAKRVVKLGAGEILLTSMDDADGT	179
		: :				: : :	
QY	486	GKGFIDLILKLISDAVNIPIVIASSGSAGVADHFSEVFNETNASALAAAGIFHRKEVPFIKAV	545	DB	180	KSGYDRLRTETEEKSVSIPVIASGCCGHTDHIIIEVFQTTVDALAASIIFYHGEATVQS	239
		: : :				: : :	
QY	546	KEHLKKEGIEVR	557	DB	240	KRKLRDANVEVR	251
		: : :				: : :	

```

RESULT 15
US-10-369-493-19698
; Sequence 19698, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 19698
; LENGTH: 256
; TYPE: PRT
; ORGANISM: Nitrosomonas europaea

```

US-10-369-493-19698

Query Match	15.7%;	Score	452.5;	DB	12;	Length	256;
Best Local Similarity	36.7%;	Pred. No.	2.5e-35;				
Matches	116;	Conservative	43;	Mismatches	94;	Indels	63;
						Gaps	8;
Qy	246	LAKRVIACLDVRANDNDGLVVTGKGQDVDRERTENEVRNLGKPVELGAGVQYVLGDGADSVS	305				
Db	1	LAKRIIPCILIK-----DGRVVKGNP-----VSLRDAGDPVEIARSTNEQAGDELV	47				
Qy	306	FLNITGTF---RDFFGLDPLQVLQRASENVFELTVGGGIRDTFDANGRYYSLEVASE	362				
Db	48	FLDITASSENRL-----ILHIVEKVAQAQVFIPLTVGGGVKKAEDVR-----R	90				
Qy	363	YFRSGADKVSIGSDAVYTAEEYIKTVKTKGSSSEQISTVYGNQAVVVSIDPRRVYLKRP	422				
Db	91	LLNAGADKKVINTSNAV-----LNPMLIKESADHYGSCQIVTAIDARQI-----	133				
Qy	423	DEVEFKAIVSHPGNGBEYAWYQCTVYNGGREGRPIGAYELAKAVEELGAGEILLNCIDC	482				
Db	134	-----PDANPESPRW-EVFTHTGGKRKPTGIDAIIEWAQKIQA LGAGEILLTSMDR	180				
Qy	483	DGQKGFDIDILKILISDAPNIPVTIASSGAGVADHFSEVFNETNSAALAAGTIFHRKEVPI	542				
Db	181	DGTRSGFDLTLTRAISDSVDLPVTIASSGVGHLDHLVEGILACHADAVLAASIFHYGVEYSI	240				
Qy	543	KAVKEHLLKKGIEVRL	558				
Db	241	LOAKOYLLSSHGIEVRL	256				

Search completed: January 20, 2004, 16:32:45
Job time : 40 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:24:22 ; Search time 21 Seconds
(without alignments) 2555.339 Million cell updates

Title: US-09-831-233A-2

Perfect score: 2877
Sequence: 1 HEKELASTKPONGFRIRAAL.....EVPKAVKEHLLKEGIEVRL 558

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 2833308

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	2284.5	79.4	592	2	T04813	amidotransferase (
2	2160	75.1	593	2	T48876	glutamine amidotra
3	1431.5	49.8	541	2	T40745	probable histidine
4	1363	47.4	552	2	S46125	amidotransferase H
5	529.5	18.4	271	2	C69352	imidazoleglycerol-
6	532	18.1	301	2	F69045	imidazoleglycerol-
7	499.5	17.4	272	2	C64351	imidazoleglycerol-
8	496.5	17.3	253	2	C70317	cyclase hisP (simi
9	481.5	16.7	253	2	B97016	imidazoleglycerol-
10	480.5	16.7	251	2	AD1504	cyclase hisP homol
11	478.5	16.6	273	2	F84242	imidazoleglycerol-
12	467.5	16.2	251	2	AD1145	cyclase hisP homol
13	467.5	16.2	267	2	D70819	cyclase hisP (simi
14	454	15.8	256	2	H83002	imidazoleglycerol-
15	450.5	15.7	252	2	A99976	cyclase-like prote
16	449.5	15.6	251	2	E90206	hypothetical prote
17	449	15.6	257	2	JE0214	cyclase hisF - Cor
18	448	15.6	253	2	C72304	cyclase hisP TM103
19	447.5	15.6	259	2	AH2167	cyclase hisP TM103
20	441	15.3	198	2	D69070	imidazoleglycerol-
21	432.5	15.0	251	2	T35077	probable cyclase h
22	432.5	15.0	252	2	B69641	cyclase hisF - Bac
23	432	15.0	208	2	A71504	amidotransferases
24	428	14.9	255	2	B81929	probable imidazole
25	426	14.8	255	2	H81176	hisF protein NMB06
26	423.5	14.7	252	2	B84097	hisF protein NMB06
27	416.5	14.5	261	2	T45251	probable cyclase h
28	410.5	14.3	261	2	F75631	cyclase hisF - Syn
29	409	14.2	208	2	A71145	amidotransferases

ALIGNMENTS

RESULT 1

REF ID: A66081

amidotransferase (EC 2.4.2.-) / cyclase - *Arabidopsis thaliana*

N;Alternate names: protein F10M23.240

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999

C;Accession: T04813

R; Bevan, M.; Lecharyn, A.; Chedfor, F.; Krivitzky, M.; Kr

submitted to the Protein Sequence Database, February 1999

A;Reference number: Z15385

A;Accession: T04813

A;Molecule type: DNA

A;Residues: 1-592 <BEV>

A; Cross-references: EMBL:AL035440

A; Experimental source: cultivar Columbia; BAC clone F10M23
C: Genotype.

C;Genetics:
A:Man posit

A:Map position: 4
A:Introns: 63/1.

A; INCRONS: 63/1; 9
A; Note: F10M23 240

A;NOLE: FUMZ3.240
C:Function:

A: Description:

A: Pathway: histidine biosynthesis

C: Keywords: glycosyltransferase;

Query Match	79.4%;	Score 2284.5;	D8 2;	Length 592;
Best Local Similarity	80.7%;	Pred. No. 2.9e-150;		
Matches 435;	Conservative 56;	Mismatches 45;	Indels 3;	Gaps 2;
Qy	21	AGAGGDSVVTLLDYGAGNVRSVRNALRTIGFDIKDVQKPEDILNAKRLIPPGVGAFAPAM	80	
Db	56	ASTSDSVVTLLDYGAGNVRSIRNALRHGFGSKDVTPGDIILNADRLIPPGVGAFAPAM	115	
Qy	81	DVLIRKGLAEALCTYIQNDRPPFLIGICLGLQLLPESSEENGPIQGLGLIPGRVGRFESSNG	140	
Db	116	DVLNRTGMAELCKYIENDRPPFLIGICLGLQLLPDSSEENGPPVKGLGVIPIGVGRFDASAG	175	
Qy	141	LRVPHIGWALDITKEGASILLDDVGNQHVYFVHSYRA-NAEDNKKEWISSTCSYGDDFTASI	199	
Db	176	IRVPHIGWALQVGKQSEILLDDVGNRHVYFVHSYRAIPSDENKDWISSTCNYESFISII	235	
Qy	200	QKGNVHAVGFHPKSGGVGLSILRRFLNADS PNKKQKPMNGKASKLAKRVIACLDVRAN	259	
Db	236	RRGNVHAVGFHPKSGGEVGLSVLRRFLHPKPL--PATQKPMEGHAKSLAKRVIACLDVRTN	293	
Qy	260	DNGDLVVTTKGDQYDVRRERTEENEVRLNKGFLVLAGQYYLDGADVEVSFLNITGRPDPLGCD	319	
Db	294	DKGDLVVTTKGDQYDVREQNEVNEVRLNKGFPVDLAGQYYKDGADIEISFLNITGRPDPLGCD	353	
Qy	320	LPMLQVLQRASENVFVPLTVGGGIRPDTANGRYYSLSLEVASVYFRSGADKVISGSDAVY	379	
Db	354	LPMIQVLRQTSKNVFPVLTGVGGIRPDTASGRYYSSLEVAAEYFRSGADKISIGSDAVS	413	

Db 410 HHVVKTSRLGPNGEAYCYOCTVKGREYRDIDVVELTRACEAGAGVILLNCMDQGSN 469
 QY 487 KGPDIIDLIKISDAVNIPVITASSGAGVADHFSVFNETNASSAALAAAGIPIHRKEVPKAYK 546
 Db 470 AGYDIELVLRLVKNVSNIPVITASSGAGIPQHFEEVFKETDCDAALAAAGIFHRQTCRIEDVK 529
 QY 547 EHLKKEGIEVR 557
 Db 530 EYLAIHDLVLR 540

RESULT 4
 S46125
 A:amidotransferase HIS7 (BC 2.4.2.-) / cyclase HIS7 - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein YBR1640; protein YBR248C
 C:Species: Saccharomyces cerevisiae
 C>Date: 26-Aug-1994 #sequence_revision 09-Sep-1994 #text_change 29-Oct-1999
 C:Accession: S46125; S46129; A48651; S35877
 R:Aljinovic, G.; Pohl, F.M.; Pohl, T.M.
 submitted to the Protein Sequence Database, August 1994
 A:Reference number: S45906
 A:Accession: S46125
 A:Molecule type: DNA
 A:Residues: 1-552 <ALJ>
 A:Cross-references: EMBL:Z36117; NID:G536662; PIDN:CAA85211.1; PID:G536663; MIPS:YBR248C
 A:Experimental source: strain S288C
 R:Rigle, M.; Baclet, M.C.; Barthe, C.; Biteau, N.; Crouzet, M.; Dolignon, P.
 submitted to the Protein Sequence Database, August 1994
 A:Reference number: S45940
 A:Accession: S46129
 A:Molecule type: DNA
 A:Residues: 1-58 <ATG>
 A:Cross-references: EMBL:Z36117; MIPS:YBR248C
 A:Experimental source: strain S288C
 R:Kuenzler, M.; Balmelli, T.; Egli, C.M.; Paravicini, G.; Braus, G.H.
 J. Bacteriol. 175, 5548-5558, 1993
 A:Title: Cloning, primary structure, and regulation of the HIS7 gene encoding a bifunctional
 A:Reference number: A48651; MUID:93374850; PMID:8366040
 A:Accession: A48651
 A:Molecule type: DNA
 A:Residues: 1-53, 'A', 55-552 <KUE>
 A:Cross-references: GB:X69815; NID:G395154; PIDN:CAA49469.1; PID:G395155
 C:Genetics:
 A:Gene: SGD:HIS7
 A:Cross-references: SGD:S0000452; MIPS:YBR248C
 A:Map position: 2R
 C:Function:
 A:Description: glycosyltransferase; pentosyltransferase
 A:Pathway: histidine biosynthesis
 A>Note: fifth and sixth step in histidine biosynthesis
 C:Keywords: glycosyltransferase; histidine biosynthesis; pentosyltransferase; transmembrane
 F:78-94/Domain: transmembrane #status predicted <TM>
 F:323-339/Domain: transmembrane #status predicted <TM>

Query Match 47.4%; Score 1363; DB 2; Length 552;
 Best Local Similarity 50.4%; Pred. No. 1.6e-86;
 Matches 281; Conservative 91; Mismatches 149; Indels 36; Gaps 11;

QY 28 VVTLLDYGAGNVRNRAIRTLGFDIKDVQKPED--ILNAKRLLIPFGVGAFAPAMDVLIR 85
 Db 3 VVHVIVDESGLQSLTNAIEHLGVEVQLVKSPPDNISGTSRLILPGVNGYHFDNLFN 62
 QY 86 KGLAEALCTVIQNDPPLGLICGLQLFPESSENGPIQGLIPGRVGRPESNGLRVPH 145
 Db 63 RGPEKPIREYESGKPTMGICVGLQALFAGSVSPKSTGLNYIDFKLSRFDSDSE-KPVPE 121
 QY 146 IGHMALDIKGSAT-LDDVGNQHVYFVHSVRA--NAB-----DNKEWISSTCSYG-DDFI 196
 Db 122 IGWNSCIPSENLPFGLDPY--KRYFVHSFAAILNSEKKKNLENLGDGKWKIAKYGSEFI 179
 QY 197 ASIQKGNVHAVQHPKESGGVGLSIILRRFLNADSFNNKQKP-----MNGKA 243

Db 180 AAVNNKNIFATQHPHPEKSGAGLANVIENFL-----KQSPPIPNYSAEKELLNDYS 232
 QY 244 S-KLAKRVITACLDVRANDGDLVVTGKQDYDVRERTEENEVRNLKRPVLAGQYLDGAD 302
 Db 233 NYGLFRIRIACLDVRTNDGDLVVTGKQDYDVRKSDGKGVNLKRPVLQAQKYQGAD 292
 QY 303 EYSLNITFRDPPGLDPLMLQVLRASENVFVPLTVGGGIRDFDTDANGRYYSLSLASE 362
 Db 293 EVTFNLNITFRDCPLKDTPLMLVLEKQAATVFPVPLTVGGGIKDIIVDVGDKIPALSVASL 352
 QY 363 YFRSGADKYSIGSDAVYTAEEYIKTCVK--TGKSSIEQISTVYGNQAVVSIOPRRVYLR 421
 Db 353 YFRSGADKYSIGTDAVYAAEKYELNGDGTSPITETISKAYGAQAQAVISVDPRVYVNS 412
 QY 422 PDEVEFKAIKVSHPGNGREYAWYQCTVNGGREGRPIGAYELAKAVEELGAGELLNCID 481
 Db 413 QADTKNKVFETYPGNGEKYCWYQCTIKGRESRDLGWELTRACEALGAGELLNCID 472
 QY 482 CDGQGGKFPDIDLIKISDAVNIPVITASSGAGVADHFSVFNETNASSAALAAAGIPIHRKEVP 541
 Db 473 KDGNSGYDLELIEHVKDAVKIPVITASSGAGVPEHPEEAPLKTADACLGAGMFRHGFT 532
 QY 542 IKAVKEHLKKEGIEVRL 558
 Db 533 VNDVKEYLLEHGLKVRM 549

RESULT 5
 C69352
 A:imidazoleglycerol-phosphate synthase (cyclase) hisF AP0819 [similarity] - Archaeoglobus
 C:Species: Archaeoglobus fulgidus
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 05-May-2000
 C:Accession: C69352
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiaich, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Woose, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: C69352
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-271 <KLE>
 A:Cross-references: GB:AE001047; GB:AE000782; NID:G2689370; PIDN:AAB90415.1; PID:G26497
 C:Superfamily: cyclase hisF

Query Match 18.4%; Score 529.5; DB 2; Length 271;
 Best Local Similarity 41.6%; Pred. No. 2.8e-29;
 Matches 131; Conservative 45; Mismatches 92; Indels 47; Gaps 8;

QY 246 LAKRVITACLDVRANDGDLVVTGKQDYDVRERTEENEVRNLKRPVLAGQYLDGAD 305
 Db 2 LAKRIIPCLDVTLL-DESEARVVGK-----VEFVNLRDAGDPVLAKRYDEGADLV 52
 QY 306 FLNITGFRDFPLGDLPLMLQVLRASENVFVPLTVGGGIRDFDTDANGRYYSLSLASEYPR 365
 Db 53 FLDTA---SPGRTMIDVIERAEQVFIPTVGGGIKSIDIN-----TILS 98
 QY 366 SGADKYSIGSDAVYTAEEYIKTCVKTSISIEQISTVYGNQAVVSIOPRRVYLRKPE- 424
 Db 99 AGADKYSINTAAVKNPE-----FVREADIFGSCIVAIIDCRNFDLSKGEY 146
 QY 425 -VEFKAIKVSHPGNGREYAWYQCTVNGGREGRPIGAYELAKAVEELGAGELLNCID 483
 Db 147 IVELE-----DGTK-AWYEVVYIGRKPVGIDAVVWAKRVBELGAGEILLTSMNRD 196
 QY 484 GQKGGFDIDLIKISDAVNIPVITASSGAGVADHFSVFNETNASSAALAAAGIPIHRKEVP 543
 Db 197 GTKGQDIPITRIKISSEVNIPVITASSGAGTKGHPYEGFVEGKADACLAASIFHREIGIR 256
 QY 544 AVKEHLKKEGIEVRL 558

	Best Local Similarity	37.5%; Pred. No.	4.3e-24;	
	Matches	117; Conservative	45; Mismatches	92; Indels
	Gaps	58; Gaps	8;	
Qy	246	LAKRVIACLDVRANDGDLVVTKGDDQYDVRTEENEVRNLKPKVELAGQVYLGDADSVS	305	
Dd	3	LAKRIIPCLDV---DNGRVV--KGVKE-----NIRDAGDPVEIARYDEOGADEIT	49	
Qy	306	FLNIITGPDFPLGLPMLQLQRASENVFVLPTVGGGIRDPDTANGRVYSLEVASVFYR	365	
Dd	50	FLDITASVD---GRDTLHTVERMASQVFIPLTVTGGVRSVQDIR-----NILN	95	
Qy	366	SGADKVSIIGSDAVYTABEYIKTGVTGKSISIEQISTYGNVAQVVSIDPRRVLYLRKPDEV	425	
Dd	96	AGADKVSINTAAVNPE-----PFGEAADFGSQCIIVAID-----	131	
Qy	426	EFKAIKVSHPNGEEYAWYOCTVNGREGRPICAYELAKAVELGAGEILLNCIDCDQG	485	
Dd	132	--AKKVSAPG---EAPRWEIFTHGGRKPTGLDAVLWAKMEDLGAGEILTLTSMDQDG	184	
Qy	486	GKGFPIDLIKLIISADVNPVIASSGAGVAHFSEVFNETNASALAAGIFHRKEVPVKAV	545	
Dd	185	KSGYDLGTVAISEAVNPVVIASCGVGNLEHLAAILGKADAVLAASIHFHGYTYTPEA	244	
Qy	546	KEHLLKEGIEVR	557	
Dd	245	KAYLASRGIVVR	256	

Qy	546	KEHLLKEGIEVR	557
		:	
Db	241	KEVIROGGIAVR	252

Search completed: January 20, 2004, 16:28:05
Job time : 22 secs

RESULT 15
A99976
C:Accession: A99976
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
R:Kuroda, M.; Ohra, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogud
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID: 21311952; PMID:11418146
A:Accession: A99976
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-252 <KUR>
A:Cross-references: GB:RA000018; PID:gl3702631; PIDN:BAB43771.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: hlsf
A:Superfamily: cyvclase hlsf

Query Match	15.7%	Score	450.5	DB 2	Length	252			
Best Local Similarity	35.9%	Pred.	No. 7.3e-24						
Matches	112	Conservative	55	Mismatches	84	Indels	61	Gaps	8
Qy	246	LAKRVIACLVDRANDGDLVTYKGDQYDVRERTTEENVRLNKLKVELAGQYYLDGAEVVS	305						
Db	2	IKKRRIIPCLDVK-----DGRVYKGIQF-----KGLRDTIGNPVDLAIYVNEAGADELV	48						
Qy	306	PLNITGRDPPLGLDPLWLQVLORASENVFPLTVGGGIRDPDTPDANGRYVSYSLVASEYFR	365						
Db	49	PLDISKTEE-----GHSMLLEVTEQASRLFIPLTVGGGIQ-----SLDDITQLLN	94						
Qy	366	SGADKVSITGSDAVYTABEYIKTVKTKGSSIEQISTVYGVNQAVVSDIPRRVYLKRPDEV	425						
Db	95	HGADKVSLSNSALKNPQ-----LIKQASDFEGRCICIAID-----	130						
Qy	426	EFKAIKVSHPGNGEEYAWYQCTVNGGREGRPIGAYELAKAVEBGLAGEILLNCIDCDGQ	485						
Db	131	-----SYYPD- -ERRAHYCCT -HGGKKMTNIKVYDWVQVEQLGAGELLVTSMGHDGM	180						
Qy	486	KGKFDIDLKLIISDAVNIPIVTASSAGVADHFSEVFNETNNSAALAGIFHRKEVPTKAV	545						
Db	181	KQGPDIIBHLAKISLVNIPIITASGGGNAOHFVLFNOTDVSAGLAASILHREDRTTVQSI	240						

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:20:32 ; Search time 18 Seconds
(without alignments)

1457.828 Million cell updates/sec

Title: US-09-831-233A-2

Perfect score: 2877

Sequence: 1 HEKELASTKPFQNGFRIRAL.....EVPKAVKEHLLKEGIEVRL 558

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2284.5	79.4	522	1 HIS5_ARATH	Q98230 arabidopsis
2	1431.5	49.8	541	1 HIS5_SCHPO	Q94303 schizosacch
3	1363	47.4	552	1 HIS5_YEAST	P33734 saccharomyc
4	1315.5	45.7	553	1 HIS5_EMENI	Q9P4P9 emericicella
5	529.5	18.4	271	1 HIS6_ARCFU	O29439 archaeoglob
6	527	18.3	272	1 HIS6_METKA	O8TYW8 methanopyru
7	522	18.1	274	1 HIS6_METH	O27398 methanobact
8	510.5	17.7	253	1 HIS6_THETN	Q8R885 thermoaer
9	499.5	17.4	272	1 HIS6_METJA	O57854 methanococc
10	496.5	17.3	253	1 HIS6_AQUAE	O66567 aquifex aeo
11	493.5	17.2	273	1 HIS6_METAC	Q8TT96 methanosarc
12	489.5	17.0	273	1 HIS6_METMA	O8PW32 methanosarc
13	481.5	16.7	253	1 HIS6_CLOAB	Q97KH8 clostridium
14	481.5	16.7	258	1 HIS6_CORGL	O31139 corynebacte
15	480.5	16.7	251	1 HIS6_LISIN	Q92E88 listeria in
16	480	16.7	254	1 HIS6_THET	Q9P4P4 thermoaer
17	478.5	16.6	273	1 HIS6_HALN1	Q9HR52 halobacteri
18	474	16.5	251	1 HIS6_PYRPU	P58800 pyrococcus
19	472	16.4	253	1 HIS6_PYRAE	Q8ZY16 pyrobaculum
20	467.5	16.2	251	1 HIS6_LISMO	Q8Y9G5 listeria mo
21	467.5	16.2	267	1 HIS6_MYCTU	O53908 mycobacteri
22	459.5	16.0	249	1 HIS6_SULTO	Q97020 sulfobolus
23	454	15.8	256	1 HIS6_PSEAE	Q9HU44 pseudomonas
24	450.5	15.7	252	1 HIS6_STPAM	Q99QW8 staphylococ
25	449.5	15.6	251	1 HIS6_SULSO	O33774 sulfobolus
26	448	15.6	253	1 HIS6_THEMA	Q9X066 thermotoga
27	447.5	15.6	259	1 HIS6_ANASP	Q8Y313 anabaena sp
28	444.5	15.5	252	1 HIS6_STAAW	O8NU13 staphylococ
29	441	15.3	198	1 HIS5_METTH	O27568 methanobact
30	432.5	15.0	251	1 HIS6_STRCO	Q9827 streptomyc
31	432.5	15.0	252	1 HIS6_BACSU	O34727 bacillus su
32	432	15.0	208	1 HIS5_LISIN	Q92E86 listeria in
33	428	14.9	255	1 HIS6_NEIMA	Q9JVB5 neisseria m

34	427.5	14.9	251	1 HIS6_CHLTE	Q8KCB0 chlorobium
35	426	14.8	255	1 HIS6_NEIMB	Q9K0H4 neisseria m
36	426	14.8	258	1 HIS6_RHIME	Q92TB3 rhizobium m
37	423.5	14.7	252	1 HIS6_BACHD	Q9K626 bacillus ha
38	421	14.6	256	1 HIS6_RALSO	Q8XW85 ralstonia s
39	416.5	14.5	261	1 HIS6_MYCLE	Q9X7C2 mycobacteri
40	410.5	14.3	261	1 HIS6_SYNY3	P74106 synchocyst
41	409	14.2	208	1 HIS5_LISMO	Q8Y9G3 listeria mo
42	409	14.2	261	1 HIS6_AZOBR	P26721 azospirillu
43	407	14.1	214	1 HIS6_BACHD	Q9K624 bacillus ha
44	406	14.1	263	1 HIS6_RHILO	Q98CT1 rhizobium l
45	405	14.1	261	1 HIS6_BRUME	Q8YE37 brucella me

ALIGNMENTS

RESULT 1					
HIS5_ARATH					
ID	HIS5_ARATH	STANDARD;	PRT;	592 AA.	
AC	Q9S230; O80330;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	imidazole glycerol phosphate synthase hisH, chloroplast precursor				
DE	(IGP synthase) (ImGP synthase) (IGPS) [Includes: Glutamine				
DE	amidotransferase (EC 2.4.2.-); Cylase (EC 4.1.3.-)]				
GN	AT4G26900 OR F10M23.240.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.				
OX	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=cv. Columbia;				
RX	MEDLINE=98316711; PubMed=9654139;				
RA	Fujimori K., Ohta D.;				
RT	"An Arabidopsis cDNA encoding a bifunctional glutamine				
RT	amidotransferase/cyclase suppresses the histidine auxotrophy of a				
RT	Saccharomyces cerevisiae his7 mutant.";				
RL	FEBS Lett. 428:229-234 (1998).				
RN	[2]				
RC	SEQUENCE FROM N.A.				
RA	STRAIN=cv. Columbia;				
RA	MEDLINE=20083488; PubMed=10617198;				
RA	Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,				
RA	Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,				
RA	Harris B., Ansoerge W., Brandt P., Grivell L.A., Rieger M.,				
RA	Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,				
RA	Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,				
RA	Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,				
RA	Vos P., Honeisel J., Zimmermann W., Wedler H., Ridley P.,				
RA	Langham S.-A., McCullagh B., Bilham L., Robben J.,				
RA	Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,				
RA	Braekens M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,				
RA	Weitzengerter T., Bothe G., Ramsperger U., Hilbert H., Braun M.,				
RA	Holzner E., Brandt A., Peters S., van Staveren M., Dirkes W.,				
RA	Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,				
RA	Bernisier S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,				
RA	De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,				
RA	Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,				
RA	Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,				
RA	Pattett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,				
RA	Borkova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,				
RA	Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,				
RA	Gabel C., Fuchs P., Fartmann B., Grandrath K., Dauner D., Herzl A.,				
RA	Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E.,				
RA	Masgenet O., Quigley F., Clabaud G., Muendlein A., Felber R.,				
RA	Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,				
RA	Chetou F., Cooke R., Berger C., Monfort A., Casacuberta E.,				
RA	Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,				
RA	Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,				

CC -1- PATHWAY: Histidine biosynthesis; fifth step.
CC -1- INDUCTION: By amino acid starvation. It has a GCN4-dependent and
CC a GCN4-independent (basal) expression.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE HISA / HISP
CC FAMILY.
CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X69815; CAA49469.1; -
DR EMBL; Z36117; CAA85211.1; -
DR PIR; S46125; S46125.
DR PDB; 1JVN; 12-OCT-01.
DR SGD; S0000452; HIS7.
DR InterPro; IPR000991; GATase_1.
DR InterPro; IPR006062; His_biosynth.
DR InterPro; IPR004651; HisF.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF00977; His_biosynth; 1.
DR TIGRFAMs; TIGR00735; hisF; 1.
DR PROSITE; PS00442; GATASE, TYPE 1; 1.
KW Histidine biosynthesis; Transferase; Glutamine amidotransferase;
KW Lyase; Multifunctional enzyme; 3D-structure.
FT DOMAIN 1 210 AMIDOTRANSFERASE.
FT DOMAIN 236 552 CYCLASE.
FT ACT_SITE 83 83 GATASE (BY SIMILARITY).
FT ACT_SITE 193 193 GATASE (BY SIMILARITY).
FT ACT_SITE 195 195 GATASE (BY SIMILARITY).
FT ACT_SITE 245 245 POTENTIAL.
FT ACT_SITE 404 404 POTENTIAL.
FT CONFLICT 54 54 G -> A (IN REF. 2).
SQ SEQUENCE 552 AA; 61068 MW; 046E11EA5F93ABA3 CRC64;

Query Match 47.4%; Score 1363; DB 1; Length 552;
Best Local Similarity 50.4%; Pred. No. 8.6e-85;
Matches 281; Conservative 91; Mismatches 149; Indels 36; Gaps 11;

Qy 28 VVTLLDYGAGNVRVNAIRTLGFDIKDVOKPED--ILNAKRLIFPGVGAFAPAMDVLI 85
Db 3 VVRVIDESGNLQSLTNAIEHLGTEVQLVKSPKDFNISGTSRLLPDGVNGYHFVNLFN 62
Qy 86 KGLAEALCTYQNDRPFLGICGLQLLSESENGPTQGLGLIPGRVGRFESSNGLRVPH 145
Db 63 RGEKPIREYESGKPTMGICVGLQALFAGSVESPKSTGLNYIDFKLSRPDDSE-KPVPE 121
Qy 146 IGHWALDIKGSAT-LDVGNGHYVFWHSYRA--NAB-----DNKEWISSTCSYG-DDFI 196
Db 122 IGWNSCIPSNLFPGLDPY--KRYVFWHSFAALNSEKKLENKNDGKIAKAKYSEEFI 179
Qy 197 ASTQKGNVHAVQHPKSGGVLSILRRFLNADSFNNKROKP-----NWGKA 243
Db 180 AAVNNKNIFATQHPKSGKAGLVNFIENFL-----KQSPPIPNVSAEKKLLMNDYS 232
Qy 244 S-KLAKRVIACLDVRANDGDLVVTQGDQYDVRERTEENVRNLGKPVLAGQYLDGAD 302
Db 233 NYGLTRRIIACLDVRTNDQGLVVTQGDQYDVRKSDGKGVNLGKPVLAQKYQGGAD 292
Qy 303 EVSFLNTGTRDPLGLDPLMLQVLRASENVFPLTVGGGIRPDTDANGRYYSLSLEAVE 362
Db 293 EVTFNLNITSFRDCPLKDTPLMLEVLKQAATVFVPLTVGGGIKOIVDVGDKI PALEVASI 352
Qy 363 YFRSGADKVSIGDVAVTAEYIKTVGK--TGKSSIEQISTVYGNQAVVVSIDPRRVYLRK 421
Db 353 YFRSGADKVSIGDVAVTAEYIKTVGK--TGKSSIEQISTVYGNQAVVVSIDPRRVYLRK 412
Qy 422 PDEVEFKAIKVSHPGPNGEYAWYQCTVNGGREGRPITGAYELAKAVBELGAGBILNCID 481

Db 413 QADTKNKVFETBYPGNGEYKWCYQCTIKGGRSRLDGLVWELTRACEALGAGBILNCID 472
Qy 482 CDGQGGKGFIDILKILSDAVNIPVIASSGAGVADHFSEVFNETNAAALAAAGIFHRKEVP 541
Db 473 KQGSNGYDLELIEHVKDAVKIPVIASSGAGVPEHFEAFELKTRADACLGAGNHRGFT 532
Qy 542 IKAVKEHLLEKGEIVRL 558
Db 533 VNDVKEYLLEHGLKVRM 549

RESULT 4
HIS5 EMENI STANDARD; PRT; 553 AA.
ID HIS5 EMENI STANDARD; PRT; 553 AA.
AC Q9P4E9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Imidazole glycerol phosphate synthase hisHF (IGP synthase) (ImGP
DE synthase) (IGPs) [includes: Glutamine amidotransferase (EC 2.4.2.-);
DE Cyclase (EC 4.1.3.-)].
GN HISHF.
OS *Emmericella nidulans* (*Aspergillus nidulans*).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; *Emmericella*.
RX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A234;
RX MEDLINE=21176419; PubMed=11277623;
RA Valerius O., Draht O., Kuebler E., Adler K., Hoffmann B., Braus G.H.;
RT "Regulation of hisHF transcription of *Aspergillus nidulans* by adenine
RT and amino acid limitation".
RL Fungal Genet. Biol. 32:21-31(2001).
CC -1- FUNCTION: IGPs catalyzes the conversion of PRFAR and glutamine to
CC IGP, AICAR and glutamate. The glutamine amidotransferase domain
CC provides the ammonia necessary to the cyclase domain to produce
CC IGP and AICAR from PRFAR (By similarity).
CC -1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulo-1-
CC ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-
CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC -1- PATHWAY: Histidine biosynthesis; fifth step.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE HISA / HISP
CC FAMILY.
CC -1- SIMILARITY: Contains 1 type-1 glutamine amidotransferase domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF159463; AAF80376.1; -
DR InterPro; IPR000991; GATase_1.
DR InterPro; IPR006062; His_biosynth.
DR InterPro; IPR004651; HisF.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF00977; His_biosynth; 1.
DR TIGRFAMs; TIGR00735; hisF; 1.
DR PROSITE; PS00442; GATASE, TYPE 1; 1.
KW Histidine biosynthesis; Transferase; Glutamine amidotransferase;
KW Lyase; Multifunctional enzyme.
FT DOMAIN 1 211 AMIDOTRANSFERASE.
FT DOMAIN 232 553 CYCLASE.
FT ACT_SITE 81 81 GATASE (BY SIMILARITY).
FT ACT_SITE 194 194 GATASE (BY SIMILARITY).
FT ACT_SITE 196 196 GATASE (BY SIMILARITY).
FT ACT_SITE 241 241 POTENTIAL.
FT ACT_SITE 403 403 POTENTIAL.
SQ SEQUENCE 553 AA; 60359 MW; 8D48C524F3058FC2 CRC64;

```
Query Match 45.7%; Score 1315.5; DB 1; Length 553;
Best Local Similarity 49.6%; Pred. No. 1.4e-81;
Matches 276; Conservative 90; Mismatches 151; Indels 39; Gaps 11;

QY 29 VTLLDYCAGNVRVRNARTLGFIDKDVQKPEIDILNAKRLIFPGVGAFAFAMDLTRKGL 88
DB 4 VHLDDYVAGNVRSLNAINKVGVEVWRSPSLKOVKEKILPVGFGHGLSGLSGGY 63
QY 89 AEALCTVIONDRPFLGICLQLLFSSSEBNGPIQGLLIPGVRGPSSNGLVRPHIGW 148
DB 64 LQPIREHAGKGFPMGICVGLQSLFSSSEEDPNIPGLTTPALRLKPDAT-KSVPHIGW 122
QY 149 H-ALDIKEGSAIILDDVGNQ-----HYVFVHSYRANAE----DNKEWISSTCSYGD 194
DB 123 NSATDTR-----IDSTGQTGYGLSPSKYVYVHSAAPVEPGILEKDGMLVARRVYGE 177
QY 195 -PTASTQKGNHVAQHPPEKSGGVLSILRFLNADS--FNNKRPQPMNGKASKLAKRVI 251
DB 178 KFIGAIARDNIFATQHPPEKSGGGRPHPSRFLGRSSAPFCHTRDSILTGEKNGLTRRII 237
QY 252 ACIDVRANDGDLVVTGDDQDVVRTEEN---EVRNLGKPVLAGQYLLDGADEVSLFN 308
DB 238 ACIDVRTNDVGLVVTGDDQDVREKDGADAGQVRLGKRPVDMAXKYFEGQADEVTFLN 297
QY 309 ITGFRDFPLGLDPLMLQVLRASENVFVPLTVGGGIRDFTDANGRYYSLSLEVASSEYFRSGA 368
DB 298 ITSFRNCLADPLMLLEIRLTSRTVFPVPLTIGGIRDVTDTGTHPALDVAWSYFKSGA 357
QY 369 DKVYSIGSDAVYTAEEYIKTG-VKTKSSIRIQISTVYGNQAVVSIIDPRVYLRKPDVEVF 427
DB 358 DKVYSIGSDAVVAEDYAAAGKVLSGKTAIETISKAYGNQAVVSDVPKRVYVSGQPDTKH 417
QY 428 KALKVSHPGNGREYAWYQCTVNGREGRPIGAYELAKAVEELGAGEILLNCIDCDGQGG 487
DB 418 RTIETFPNNAAGNFCWYQCTIIGKRETRDLQVQVAVMAGAGEILLNCIDKDGNS 477
QY 488 GFDIDLKILSD-----AVNIPVIASSGAGVADHFSEVFNETNAGAAALAGIFHRKEVP 541
DB 478 G-----SILTDQPPSKRAVKIPVIASSGAGMPGHFEVEVDQITTDALGAGMFHREY 531
QY 542 IKAVKHELLKEGIEVR 557
DB 532 VGEVQKQYLEDGFLVR 547

RESULT 5
HIS6_ARCFU STANDARD; PRT; 271 AA.
ID HIS6_ARCFU
AC 029439;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP
DE synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase
DE subunit hisF) (IGPs subunit hisF).
GN HISF OR AF0819
OS Archaeoglobus fulgidus.
OC Archaeaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MDLLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-P., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.P., McDonald L., Uterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
```

```
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -I- FUNCTION: IGPs catalyzes the conversion of PRFAR and glutamine to
CC IGP, AICAR and glutamate. The hisF subunit catalyzes the
CC cyclization activity that produces IGP and AICAR from PRFAR using
CC the ammonia provided by the hisH subunit (By similarity).
CC -I- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-
CC ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-
CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC aminimidazole-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC -I- PATHWAY: Histidine biosynthesis; fifth step.
CC -I- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AS001047; AAB90415.1; -.
CC PIR; C69352; C69352.
CC TIGR; AF0819; -.
CC HAMAP; MF_01013; -.
CC InterPro; IPR003009; FMN enzyme.
CC InterPro; IPR006062; His biosynth.
CC InterPro; IPR004651; HisF.
CC Pfam; PF00977; His biosynth; 1.
CC TIGRFAMs; TIGR00735; hisF; 1.
CC Histidine biosynthesis; Lyase; Complete proteome.
KW ACT SITE 11 11
FT ACT SITE 134 134 POTENTIAL.
FT ACT SITE 134 134
SQ SEQUENCE 271 AA; 29861 MW; 91A5EA022AA3A5DA CRC64;

Query Match 18.4%; Score 529.5; DB 1; Length 271;
Best Local Similarity 41.6%; Pred. No. 7.5e-29;
Matches 131; Conservative 45; Mismatches 92; Indels 47; Gaps 8;

QY 246 LAKRVIACLDVRANDGDLVVTGDDQDVVRTEENENLGNKPKVELAGQYLLDGADEV 305
DB 2 LAKRIIPCLDVTI--DESEARVVRG-----VEFVNLRDAGDPFELAKRYDEEGDELV 52
QY 306 FLNITGFRDFPLGLDPLMLQVLRASENVFVPLTVGGGIRDFTDANGRYYSLSLEVASSEYFR 365
DB 53 FLDTIA---SPGRRTMIDVIERTAEOVFIPFTVGGGKISIEDIN-----TILS 98
QY 366 SGADKVSIGSDAVYTAEEYIKTGKTKGKSSISQISTVYGNQAVVSIIDPRVYLRKPD- 424
DB 99 AGADKVSINTAAVKNPE-----FVREAADIFGQCITVAIDCRRNFDLSKGEY 146
QY 425 -VEFKAIKVSHPGNGEEYAWYQCTVNGREGRPICAYELAKAVEELGAGEILLNCIDCD 483
DB 147 IVELE-----DGTK-ANWEVYIYGRKPGVIDAVVMWAKRVEELGAGEILLTSMNRD 196
QY 484 GQKGFIDILIKLISDANIPVITASSGAGVADHFSEVFNETNAGAAALAGIFHRKEVP 543
DB 197 GTYDGFIDITRKISEEVNIPVIASSGAGTKHEFEGFVEGKADACLAASIFHYREIGR 256
QY 544 AVKHELLKEGIEVRL 558
DB 257 EIKYLAERGQVRL 271

RESULT 6
HIS6_METKA
ID HIS6_METKA STANDARD; PRT; 272 AA.
```

Q8TYW8;
28-FEB-2003 (Rel. 41, Created)
28-FEB-2003 (Rel. 41, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP
DE synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase
DE subunit hisF) (IGPS subunit hisF).
GN HSF OR MK0173.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezhevaeva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozhavkin S.A.,
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC -1- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to
CC IGP, AICAR and glutamate. The hisF subunit catalyzes the
CC cyclization activity that produces IGP and AICAR from PRFAR using
CC the ammonia provided by the hisH subunit (By similarity).
CC -1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-
CC ylaminomethylideneamino)-1-(5-phosphoribosyl)imidazole-4-
CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC -1- PATHWAY: Histidine biosynthesis; fifth step.
CC -1- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; AE010316; ANM01390.1; -
CC HAMAP; MF_01013; -; 1.
CC InterPro; IPR003009; FMN enzyme.
CC InterPro; IPR006062; His biosynth.
CC InterPro; IPR004651; HisF.
CC Pfam; PF00977; His biosynth; 1.
CC TIGRfams; TIGR00735; hisF; 1.
CC Histidine biosynthesis; lysase; Complete proteome.
FT ACT SITE 12 12 POTENTIAL.
FT ACT SITE 131 131 POTENTIAL.
SQ SEQUENCE 272 AA; 29848 MW; CA7F9018E3A1A0CC CRC64;

Query Match 18.3%; Score 527; DB 1; Length 272;
Best Local Similarity 41.5%; Pred. No. 1.le-28;
Matches 131; Conservative 42; Mismatches 91; Indels 52; Gaps 8;

QY 246 LAKRVIACLDVRANDGDLVVTGKDQYDVRTEENRNLGKVELAGQVYLDGADEV 305
DB 3 LAKRIIFCLDVK-----DGRVVKGRF-----RGLRDAGDPAELAHYHGHGADEIV 49
QY 306 FLNITGFRDFPLGLDPLQLQRASENVFPLTVGGGIRGFTDANGRYYSLEVASFYR 365
DB 50 FLDISA---SPEGRLMVDVVRRTAEKVFIPMTVGGGISDVED-----FR 91
QY 366 -----SGADKVSIGSDAVYTAEEYIKTVGKTKSSIEQISTVYQNAVVSIDPRVYLK 421
DB 92 RALTAGADKRVSVNTAAENPE-----LISEAADIFGQCQVVAIDAKREPL-K 138
QY 422 PDEVEFKAIVKSHPGNGEYAWYQCTVNGGREGRPITGAVELAKAKAEEELGAGELLNCID 481

DB 139 PEHEHV-----ADHIFSNDDGEYWFVYVRGREGPVDLDAITWAKRYVEELGAGILLTSD 194
QY 482 CDQGGKGFDDIKLISDAVNIPVIASSGAGVADHSEVFNETNAGSALAAGIFFHKEVP 541
DB 195 ADGTQEGYDIELTREVCAVSPVIASGGCGHPKHWVFKEADADALAASIFHYGKFT 254
QY 542 IKAVKEHLKKEGIEVR 557
DB 255 IEEVKEHLAERGVRVR 270

RESULT 7
HIS6 METTH STANDARD; PRT; 274 AA.
AC C27356;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP
DE synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase
DE subunit hisF) (IGPS subunit hisF).
GN HSF OR MTH1343.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwan N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7155-7155(1997).
CC -1- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to
CC IGP, AICAR and glutamate. The hisF subunit catalyzes the
CC cyclization activity that produces IGP and AICAR from PRFAR using
CC the ammonia provided by the hisH subunit (By similarity).
CC -1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-
CC ylaminomethylideneamino)-1-(5-phosphoribosyl)imidazole-4-
CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC -1- PATHWAY: Histidine biosynthesis; fifth step.
CC -1- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; AE000897; AAB85821.1; ALT_INIT.
CC HAMAP; MF_01013; -; 1.
CC InterPro; IPR003009; FMN enzyme.
CC InterPro; IPR006062; His biosynth.
CC InterPro; IPR004651; HisF.
CC Pfam; PF00977; His biosynth; 1.
CC TIGRfams; TIGR00735; hisF; 1.
CC Histidine biosynthesis; lysase; Complete proteome.
FT ACT SITE 11 11 POTENTIAL.
FT ACT SITE 134 134 POTENTIAL.
SQ SEQUENCE 274 AA; 30463 MW; B80082BE452AC53 CRC64;


```
Query Match      18.1%; Score 522; DB 1; Length 274;
Best Local Similarity 40.6%; Pred. No. 2.5e-28;
Matches 127; Conservative 44; Mismatches 102; Indels 40; Gaps 7;

QY 246 LAKRVITACLDVRANDGDLVVTGGDQYDVRTEENEVRNLKPVLAGQYVLDGADSVS 305
DB 2 LAKRIIPCLDVK-----DLQVNG-----RVVKGVEFKQIRYAGDVELATRYEDGADIV 52

QY 306 FLNITGFRDPLGLDPLMLQVLRASENVFPLTVGGGIRDTDANGRYYSLSLEVASEYPR 365
DB 53 FLDTITASHE---RRETHVIEATENVFPCVGGGIRKPED-----YFKWL-----K 98

QY 366 SGADKVSIGSDAVYTAEEYIKTVGKSSIEQISTVYGNQAVVVSIDPRVYLRKPDEV 425
DB 99 AGADKCSNTAAIKNPE-----LINEASDLVGSQACVVAIDAKRRYIENPRES 146

QY 426 EFKAIVKSHPGNGEYAWYQCTVNGGREGRPITGAYELAKAVEELGAGEILLNCIDCDGQ 485
DB 147 DERFIIIEVDG-----YCWYECISYGGREFTGIDAVKMAWECQDRGAGEILLTSMRDGT 201

QY 486 KGKFDIDLKILSDAVNIPIASSGAGVADHSEVENETNASAAAGIFHRKEVPIKAV 545
DB 202 KMGTDIPLTMTGENDLPIVASSGGVGEPEHIYEAFDTGKADAAALASIFHFEYVPPAV 261

QY 546 KEHLKKEGIEVRL 558
DB 262 KEVLRSGVPIRL 274

RESULT 8
HIS6 THETN      STANDARD;      PRT;      253 AA.
AC QSR85;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP
DE synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase
DE subunit hisF) (IGPS subunit hisF).
GN HISF OR TTE2133.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae;
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=2192816; PubMed=1197336;
RA Bao O., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).

CC -1- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to
CC IGP, AICAR and glutamate. The hisF subunit catalyzes the
CC cyclization activity that produces IGP and AICAR from PRFAR using
CC the ammonia provided by the hisH subunit (By similarity).
CC -1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribose-1-
CC ylaminomethylideneamino)-1-(5-phosphoribosyl)imidazole-4-
CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC aminimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC -1- PATHWAY: Histidine biosynthesis; fifth step.
CC -1- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/)
```

```
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE013160; AAM25298.1; -.
DR HAMAP; MF_01013; -. 1.
DR InterPro; IPR003009; FMN enzyme.
DR InterPro; IPR006082; His_biosynth.
DR InterPro; IPR004651; HisF.
DR Pfam; PF00977; His_biosynth; 1.
DR TIGRPFAM; TIGR00735; hisF; 1.
KW Histidine biosynthesis; Lyase; Complete proteome.
FT ACT_SITE 11 11 POTENTIAL.
FT ACT_SITE 130 130 POTENTIAL.
SQ SEQUENCE 253 AA; 27516 MW; 4CAA77896FA57C2B CRC64;

Query Match      17.7%; Score 510.5; DB 1; Length 253;
Best Local Similarity 41.2%; Pred. No. 1.3e-27;
Matches 129; Conservative 42; Mismatches 81; Indels 61; Gaps 7;

QY 246 LAKRVITACLDVRANDGDLVVTGGDQYDVRTEENEVRNLKPVLAGQYVLDGADSVS 305
DB 2 LAKRIIPCLDVK-----DGRVVKGINF-----VNLKADAGDPVEIAERYNELGADLV 48

QY 306 FLNITGFRDPLGLDPLMLQVLRASENVFPLTVGGGIRDTDANGRYYSLSLEVASEYPR 365
DB 49 FLDTITASYE---XKIMIDIVKRTSEKVFIPLTUVGGGISDIDIR-----EVLK 94

QY 366 SGADKVSIGSDAVYTAEEYIKTVGKSSIEQISTVYGNQAVVVSIDPRVYLRKPDEV 425
DB 95 AGADKVSINTQAV-----KQPTLIHQALRFSGQCVVVAIDAKK-----133

QY 426 EFKAIVKSHPGNGEYAWYQCTVNGGREGRPITGAYELAKAVEELGAGEILLNCIDCDGQ 485
DB 134 -----RPDGTGYNVY---INGRINTGLDAVEWAKVKVDLGAAGEILLTSMRDGT 180

QY 486 KGKFDIDLKILSDAVNIPIASSGAGVADHSEVENETNASAAAGIFHRKEVPIKAV 545
DB 181 KQGYDIELTRLISEAVSIPVIASSGGAGKPEHFKVFTQKADAAALASVPHYGELDIKEL 240

QY 546 KEHLKKEGIEVRL 558
DB 241 KRYLKDEGIPVRL 253

RESULT 9
HIS6 METJA      STANDARD;      PRT;      272 AA.
AC Q57854;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP
DE synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase
DE subunit hisF) (IGFS subunit hisF).
GN HISF OR MJ0411.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kleravage A.R., Dougherty B.A., Tomb J.-P., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock J.F., Weidman J.F., Nguyen D.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Sadow P.W., Hanna M.C.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Cotton M.D., Roberts K.M., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
```

CC -!- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to
 CC IGP, AICAR and glutamate. The hisf subunit catalyzes the
 CC cyclization activity that produces IGP and AICAR from PRFAR using
 CC the ammonia provided by the his subunit (By similarity).
 CC -!- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulos-1-
 CC ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-
 CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
 CC aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
 CC -!- PATHWAY: Histidine biosynthesis; fifth step.
 CC -!- SUBUNIT: Heterodimer of hish and hisf (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U67493; AAB98400.1; -.
 CC DR PIR: C64351; C64351.
 CC DR TIGR: MJ0411; -.
 CC DR HAMAP: MF_01013; -; 1.
 CC DR InterPro: IPR003009; FNM enzyme.
 CC DR InterPro: IPR006062; His_biosynth.
 CC DR InterPro: IPR004651; Hisf.
 CC DR Pfam: PF00977; His_biosynth; 1.
 CC DR TIGRFAMS: TIGR00735; hisf; 1.
 CC KW Histidine biosynthesis; Lyase; Complete proteome.
 CC FT ACT_SITE 11 11 POTENTIAL.
 CC FT ACT_SITE 130 130 POTENTIAL.
 CC SQ SEQUENCE 272 AA; 30167 MW; 84943CA540A0FEBC CRC64;
 CC
 CC Query Match 17.4%; Score 499.5; DB 1; Length 272;
 CC Best Local Similarity 40.1%; Pred. No. 8.1e-27;
 CC Matches 128; Conservative 47; Mismatches 89; Indels 55; Gaps 10;
 CC
 CC QY 246 LAKRVIACLDVRANDGLVVTGQDYDVRERTEENRNLGKPVLAGQYILD-GADEV 304
 CC Db LTRKRIIPCLDIK-----DGRVVKG-----TKFLNLRDAGDPVELA-QYDDEGADEL 47
 CC
 CC QY 305 SFLNITGF---RDFPLGDLPMQLQORASENVFPLTVGGGIRFTDANGRYVSSLEVAS 361
 CC Db VFLDITASAEKRFI-----IIDVVERTAEKVFPLTVGGGIRK-----SIEDFR 90
 CC
 CC QY 362 EYFRSGADKVSIGSDAVVTAEEYIKTGKVTGKSSIEQISTVYGNQAVVVSIDPRRVYLK 421
 CC Db RILRAGADKVSINTAAV-----KNPNLKEASEIFGSCVVVAIDAKRHVNE 138
 CC
 CC QY 422 P--DEVEFKAIKVSHPGNGEYAWYQCTVNGREGRPICAYELAKAVEBELGAGEILLNC 479
 CC Db DEIDKINKNVKVE-----DGVCWFEVYIGGRKETGIDAINWAKKVEELGAGEILLTS 192
 CC
 CC QY 480 IDCQGGKGFIDILKILISDAVNIPVTASSGAGVADHFSEVFNETNASALAAGIFRKE 539
 CC Db IDKDGKSGYDLILITKEISKSVKLPVTASSGCGKPEHVFYAFVYGKADALMAGILHYRE 252
 CC
 CC QY 540 VPIKAVKEHLLKEGIEVRL 558
 CC Db YTIEEIKKYCADRGIPNRL 271

RESULT 10

HIS6 AQUAE

ID HIS6 AQUAE STANDARD; PRT; 253 AA.

AC O66567;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Imidazole glycerol phosphate synthase subunit hisf (BC 4.1.3.-) (IGP

DE synthase cyclase subunit) (IGP synthase subunit hisf) (ImGP synthase

DE subunit hisf) (IGPS subunit hisf).
 GN HISF OR AQ_181.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
 OX NCBI_TaxID=63363;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=VF5;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., R.,
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus".
 RL Nature 392:353-358(1998).
 CC -!- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to
 CC IGP, AICAR and glutamate. The hisf subunit catalyzes the
 CC cyclization activity that produces IGP and AICAR from PRFAR using
 CC the ammonia provided by the his subunit (By similarity).
 CC -!- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulos-1-
 CC ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-
 CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
 CC aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
 CC -!- PATHWAY: Histidine biosynthesis; fifth step.
 CC -!- SUBUNIT: Heterodimer of hish and hisf (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: A3000677; AAC06520.1; -.
 CC DR PIR: C70317; C70317.
 CC DR HAMAP: MF_01013; -; 1.
 CC DR InterPro: IPR003009; FNM enzyme.
 CC DR InterPro: IPR006062; His_biosynth.
 CC DR InterPro: IPR004651; Hisf.
 CC DR Pfam: PF00977; His_biosynth; 1.
 CC DR TIGRFAMS: TIGR00735; hisf; 1.
 CC KW Histidine biosynthesis; Lyase; Complete proteome.
 CC FT ACT_SITE 11 11 POTENTIAL.
 CC FT ACT_SITE 130 130 POTENTIAL.
 CC SQ SEQUENCE 253 AA; 27796 MW; E0554ESF21597556 CRC64;
 CC
 CC Query Match 17.3%; Score 496.5; DB 1; Length 253;
 CC Best Local Similarity 40.6%; Pred. No. 1.2e-26;
 CC Matches 128; Conservative 44; Mismatches 74; Indels 69; Gaps 10;
 CC
 CC QY 246 LAKRVIACLDVRANDGLVVTGQDYDVRERTEENRNLGKPVLAGQYILDGADEV 305
 CC Db LAKRIIPCLDV---DKGRVV---KGKVF-----LNLRDAGDPVEVAKRYEEGADLV 48
 CC
 CC QY 306 FLNITGF---RDFPLGDLPMQLQORASENVFPLTVGGGIRFTDANGRYVSSLEVAS 362
 CC Db VFLDITASAEKRFI-----IIDVVERTAEKVFPLTVGGGIR-----SLEDMMR 91
 CC
 CC QY 363 YFRSGADKVSIGSDAVVTAEEYIKTGKVTGKSSIEQISTVYGNQAVVVSIDPRRVYLK 422
 CC Db LLEAGADKVSINTAAVKN--PNLIYEAKR-----FGQCIVVAIDAKR----- 133
 CC
 CC QY 423 DEVEFKAIKVSHPGNGEYAWYQCTVNGREGRPICAYELAKAVEBELGAGEILLNCIDC 482
 CC Db L34 -----KGNSEVY-----IHGRTPTGLDAVEWAKKVESLGAAGEILLTSMR 175
 CC
 CC QY 483 DGCGKGFIDILKILISDAVNIPVTASSGAGVADHFSEVFNETNASALAAGIFRKEVPI 542
 CC Db DGTGKGYDIELNRAISEAVNIPVTASSGAGKKEHFEVFKTKEVAAALAAVPHFREISI 235

QY 543 KAVKEHLKEGIEVR 557
 Db 236 PELKEYLLERGINVR 250

Matches 122; Conservative 48; Mismatches 100; Indels 45; Gaps 7;

QY 246 LAKRVIAICLDVRANDNGDLVVTGKQDYVRETEENEVRNKGKPVELAQYILDGADEVS 305
 Db 2 LYKRIIPCLDVLDRAGGCVKGVFVDLKE-----AGDPVELAKRYNEDGADELV 52

QY 306 FLNITGFRDFPLGDLPLQVLRASENVFPLVTGGGIRDFDTDANGRYYSSELEVAASEYR 365
 Db 53 FLDITASAH---GRETMIDVIERTADVEPIPLTVGGGI-----SSDAIRQILR 98

QY 366 SGADKVSIGSDAVYTAAEYIKTVKTKSSISQISTVYGNQAVVVISIDPRR--VYLRKPD 423
 Db 99 AGADKRVSVNTSAVKNPD-----FIKSSDIFGAQCVITADCRNTDIKNRPD 146

QY 424 EYEFKRAIKVSHPGPNGEYAYQCTVNGREGRPICAYELAKAVELGAGEILLNCIDCD 483
 Db 147 KT---VLELEDGTP-----AYEVVYGGREATGIDAVQWAKKAEILTSMDRD 198

QY 484 GQKGFDPIDLIKIDAVNIPVASSGAGVADHFSEVFNETNASALAAAGIFHRKEVPK 543
 Db 199 GTCAGYDLPIITRKLSELDPIITASGGVGNPQHIVEGSEGRADALAAASIFHFSEYSIW 258

QY 544 AVKEHLKEGIEVRL 558
 Db 259 EVKYLREIREIPVRL 273

RESULT 12
 HIS6_METMA STANDARD; PRT; 273 AA.
 AC Q8PM92;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase subunit hisF) (IGPS subunit hisF).
 DE HISF OR MM1704.
 OS Methanosarcina mazei (Methanosarcina frigida).
 OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
 OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2209;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Goel / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
 RX MEDLINE=22120827; PubMed=12125824;
 RA Deppenmeier U., Johann A., Hartech T., Merkl R., Schmitz R.A., Martinez-Arias R., Henne A., Wierzer A., Baeumer S., Jacobi C., Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S., Bhattacharya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P., Fritz H.-J., Gottschalk G.;
 RA "The genome of Methanosarcina mazei: evidence for lateral gene transfer between Bacteria and Archaea.";
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
 CC -1- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to IGP, AICAR and glutamate. The hisF subunit catalyzes the cyclization activity that produces IGP and AICAR from PRFAR using the ammonia provided by the hisH subunit (By similarity).
 CC -1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O).
 CC -1- PATHWAY: Histidine biosynthesis; fifth step.
 CC -1- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE010714; AAM03985.1; .
 DR HAMAP; MF 01013; . 1.
 DR InterPro; IPR003009; FMN enzyme.
 DR InterPro; IPR006062; His_biosynth.
 DR InterPro; IPR004651; HisF.
 DR Pfam; PF00977; His_biosynth; 1.
 DR TIGRFAMs; TIGR00735; hisF; 1.
 FT ACT_SITE 11 11 POTENTIAL.
 FT ACT_SITE 134 134 POTENTIAL.
 SQ SEQUENCE 273 AA; 29767 MW; B3BAED4D8602C06A CRC64;
 Query Match 17.2%; Score 493.5; DB 1; Length 273;
 Best Local Similarity 38.7%; Pred. No. 2.1e-26;

```
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; AE013407; AA031400.1; -.
CC HAMAP; MF 01013; -. 1.
CC InterPro; IPR003009; FMN enzyme.
CC InterPro; IPR006062; His biosynth.
CC InterPro; IPR004651; HisF.
CC Pfam; PF00977; His biosynth; 1.
CC TIGRFAMs; TIGR00735; hisF; 1.
KW Histidine biosynthesis; Lyase; Complete proteome.
FT ACT_SITE 11 11 POTENTIAL.
FT ACT_SITE 134 134 POTENTIAL.
SQ SEQUENCE 273 AA; 29690 MW; FCC6AD1B6377918 CRC64;

Query Match 17.0%; Score 489.5; DB 1; Length 273;
Best Local Similarity 38.7%; Pred. No. 3.9e-26;
Matches 121; Conservative 48; Mismatches 103; Indels 41; Gaps 6;

QY 246 LAKRVIACLDVRANDGDLVVTGDDYDVRTEENEVRNLGKPVLAGQYLDGADSVS 305
DB 2 LTKRIIPCLDVTIDRAGGCVKGVFVDLKE-----AGDPVELAKRYNEDGADLV 52
QY 306 FLNITGPRDFPLGDLPLQVLRASENVFPLTVGGGIRDTDANGRYYSLSLEVASVEYFR 365
DB 53 FLDITASAH---GRETMIDVIETADEVFPLTVGGGI-----SSIEAIRQLR 98
QY 366 SGADKVSIGSDAVVTAEYIKTGVTGKSSIEQISTVYGNQAVVSDIPRRVLRKPDEV 425
DB 99 AGADKVSIVNTSAVKNP-----FIKSSDIFGAQCIIVTADCKR-NTNVKNDP 145
QY 426 EFPAIKVSHPGNGEYAWYQCTVNGGREGRPICAYELAKAVELGAGEILLNCIDCDGQ 485
DB 146 DKTILELEDGTP-----AWTEVYIGRKNTGIDAVQWAKRAELGSGEILLTSMRDGT 200
QY 486 GKGFIDILIKLISDAVNIPIVASSGAGVADHFESEVFNETNASAAAGIFHRKEVPIKAV 545
DB 201 CAGYDIPITRKLSELDIPIASGVGNPQHIVYGFSDGKADAAALAAASIFHFGESYQEV 260
QY 546 KEHLKKEGIEVRL 558
DB 261 KEFLKPKIPVRL 273

RESULT 13
HIS6_CLOAB STANDARD; PRT; 253 AA.
AC Q97KH8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase subunit hisF) (IGPS subunit hisF).
GN HISF OR CAC0941.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q., Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I., Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J., Bennett G.N., Koonin E.V., Smith D.R.;
RA "Genome sequence and comparative analysis of the solvent-producing bacterium Clostridium acetobutylicum.";
RT J. Bacteriol. 183:4823-4838 (2001).
CC -1- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to IGP, AICAR and glutamate. The hisF subunit catalyzes the cyclization activity that produces IGP and AICAR from PRFAR using the ammonia provided by the hisH subunit (By similarity).
```

OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13059 / AS019;
RA Jung S.I., Kwon J.H., Lee M.-S.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of *Corynebacterium glutamicum* ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to
CC IGP, AICAR and glutamate. The hisF subunit catalyzes the
CC cyclization activity that produces IGP and AICAR from PRFAR using
CC the ammonia provided by the hisH subunit (By similarity).
CC -1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-
CC ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-
CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC -1- PATHWAY: Heterodimer of hisH and hisF (By similarity).
CC -1- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF030405; AB84280.1; -
CC EMBL; AP005280; BAB99487.1; -
CC PIR; JE0214; JE0214.
CC DR HAMAP; MF 01013; -; 1.
CC DR InterPro; IPR003009; FMN enzyme.
CC DR InterPro; IPR006062; His_biosynth.
CC DR InterPro; IPR004651; HisF.
CC DR Pfam; PF00977; His_biosynth; 1.
CC DR TIGRFAMs; TIGR00735; hisf; 1.
CC FT ACT SITE 12 12 POTENTIAL.
CC FT ACT SITE 131 131 POTENTIAL.
CC FT CONFLICT 28 31 RDAG -> AMLA (IN REF. 1).
CC FT CONFLICT 118 118 S -> W (IN REF. 1).
CC FT CONFLICT 208 208 MISSING (IN REF. 1).
CC FT CONFLICT 224 224 A -> R (IN REF. 1).
CC SQ SEQUENCE 258 AA; 27245 MW; 7A28B377697B52DF CRC64;
Query Match 16.78; Score 481.5; DB 1; Length 258;
Best Local Similarity 39.48; Pred. No. 1.2e-25;
Matches 123; Conservative 49; Mismatches 83; Indels 57; Gaps 10;
QY 246 LAKRVIACLDVRANDGDLVTKGDQYDVRTEENEVRMLKPEVLAGQVYLDGADEVS 305
DB 3 VAIRVIFCLDV---DNGRVV--KGVNFE-----NLRDAGDPVELAKYDEGADEL 49
QY 306 FLNITGRDPLGLDPLMLQLORASENVFPLTVGGGIRPDTDANGRYYSLEVASFYR 365
DB 50 FLDVDTSKSH--GRGTMLDVVRRTADQVFPLTVGGGV-----SEEDVDQLLR 95
QY 366 SGADKVISGDVAVTAEYIKTGKTSKSEIEQISTVYGNQAVVSDPRVYLRKPEV 425
DB 96 AGADKVSNTSARPE-----LLSELSKRFGAQCVILSVDAARRV-----PEG- 138
QY 426 EFKAKVSHPGNCEEYAVYQCTVNGGREGRPICAVELAKAVELGAGEILLNCIDCDGQ 485
DB 139 -----GTQPSG-----FEVTHGSGKGAELDAIEWAKRGEILGVEILLNSMDGDT 186
QY 486 KGKFPDILIKLISDAPNIPVIASGAGVADHFSEVFNETNASALAAAGIFHRKEVPTKAV 545
|||||: : : : : ||||| ||||| : : : : : ||||| ||||| : : : : : |||||

DB 187 KNGFDLELLEKVRRAVSIPIVIASGGAGKAEHPPAV-AAGANAVLAATIPHFREVITAEV 245
QY 546 KEHLKKEGIEVR 557
DB 246 KGAIKDAGFEVR 257
RESULT 15
HIS6_LISIN
ID HIS6_LISIN STANDARD; PRT; 251 AA.
AC Q92E58; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Imidazole glycerol phosphate synthase subunit hisF (EC 4.1.3.-) (IGP
DE synthase cyclase subunit) (IGP synthase subunit hisF) (ImGP synthase
DE subunit hisF) (IGPS subunit hisF).
DE HISF OR LI0572.
GN Listeria innocua.
OS Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OC NCBI_TaxID=1642;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian L.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordieck G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of *Listeria* species.";
RL Science 294:849-852 (2001).
CC -1- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to
CC IGP, AICAR and glutamate. The hisF subunit catalyzes the
CC cyclization activity that produces IGP and AICAR from PRFAR using
CC the ammonia provided by the hisH subunit (By similarity).
CC -1- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulose-1-
CC ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-
CC carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC -1- PATHWAY: Histidine biosynthesis; fifth step.
CC -1- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL596165; CAC95804.1; -
CC PIR; AD1504; AD1504.
CC LITLIT; LIN00572; -
CC HAMAP; MF 01013; -; 1.
CC InterPro; IPR003009; FMN enzyme.
CC InterPro; IPR006062; His_biosynth.
CC InterPro; IPR004651; HisF.
CC Pfam; PF00977; His_biosynth; 1.
CC TIGRFAMs; TIGR00735; hisf; 1.
CC Histidine biosynthesis; Lyase; Complete proteome.
CC ACT_SITE 11 11 POTENTIAL.
CC ACT_SITE 130 130 POTENTIAL.
CC SQ SEQUENCE 251 AA; 26670 MW; F7C08CEDEA909170 CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:23:47 ; Search time 40 seconds
(without alignments)

3599.834 Million cell updates/sec

Title: US-09-831-233A-2

Perfect score: 2877

Sequence: 1 HEKELASTKQNGFRIRAAAL.....EVIKAVKEHLLKEGIEVRL 558

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2270.5	78.9	592	10 Q8GX14	Q8GX14 arabidopsis
2	1192.5	41.4	627	3 Q9HFV7	Q9HFV7 candida alb
3	484.5	16.8	258	16 Q8FNZ9	Q8FNZ9 corynebacte
4	476	16.5	261	2 Q8DQ06	Q8DQ06 heliobacill
5	439	15.3	256	16 Q8G6F7	Q8G6F7 bifidobacte
6	438	15.2	203	16 Q8R883	Q8R883 thermoanaer
7	432.5	15.0	254	16 Q8ESS2	Q8ESS2 oceanobacil
8	424.5	14.8	252	16 Q8CQ92	Q8CQ92 staphylococ
9	421	14.6	251	16 Q8DTR3	Q8DTR3 streptococ
10	412.5	14.3	281	16 Q8DJN7	Q8DJN7 synchococ
11	407	14.1	201	16 Q8KF56	Q8KF56 chlorobium
12	405	14.1	261	16 Q8FY07	Q8FY07 bruceella su
13	396	13.8	201	16 Q8ESS0	Q8ESS0 oceanobacil
14	394	13.7	201	2 Q939J6	Q939J6 campylobact
15	393	13.7	202	17 Q8TS91	Q8TS91 methanosarc
16	390.5	13.6	209	16 Q8DIP5	Q8DIP5 synchococ

17	381	13.2	207	17 Q8PVD5	Q8PVD5 methanosarc
18	378.5	13.2	257	16 Q8D8Q5	Q8D8Q5 vibrio vuln
19	378	13.1	201	16 Q8DTR1	Q8DTR1 streptococ
20	358.5	12.5	258	16 Q8FG48	Q8FG48 escherichia
21	357.5	12.4	257	16 Q8CX42	Q8CX42 shewanella
22	354	12.3	192	16 Q8CTV0	Q8CTV0 staphylococ
23	344	12.0	200	16 Q8PLG8	Q8PLG8 xanthomonas
24	343.5	11.9	204	16 Q8D8Q3	Q8D8Q3 vibrio vuln
25	339	11.8	200	16 Q8P9P1	Q8P9P1 xanthomonas
26	336.5	11.7	215	16 Q8G4S6	Q8G4S6 bifidobacte
27	335.5	11.7	221	16 Q8EFB4	Q8EFB4 shewanella
28	328	11.4	216	16 Q8FY09	Q8FY09 bruceella su
29	319	11.1	211	16 Q8FNZ6	Q8FNZ6 corynebacte
30	317.5	11.0	204	2 Q8KIU3	Q8KIU3 pseudomonas
31	313.5	10.9	200	17 Q8TV83	Q8TV83 methanopyru
32	265	9.2	251	2 Q8KIS2	Q8KIS2 pseudomonas
33	216	7.5	104	2 Q8RPQ6	Q8RPQ6 thermoanaer
34	213	7.4	166	2 Q8VNA3	Q8VNA3 rhizobium e
35	208.5	7.2	260	2 Q8KIU2	Q8KIU2 pseudomonas
36	170.5	5.9	240	16 Q8ESS1	Q8ESS1 oceanobacil
37	151	5.2	114	2 Q9F8I8	Q9F8I8 carboxydoth
38	148	5.1	502	16 Q8RJ20	Q8RJ20 streptomyc
39	145	5.0	479	2 Q939K5	Q939K5 propionibac
40	139	4.8	816	10 Q9LJG4	Q9LJG4 arabidopsis
41	138	4.8	808	10 Q94A27	Q94A27 arabidopsis
42	130.5	4.5	246	16 Q8FG49	Q8FG49 escherichia
43	129.5	4.5	234	16 Q8CQ93	Q8CQ93 staphylococ
44	128	4.4	239	16 Q8DTR2	Q8DTR2 streptococ
45	127.5	4.4	937	17 Q8PTT2	Q8PTT2 methanosarc

ALIGNMENTS

RESULT 1

Q8GX14 ID Q8GX14 PRELIMINARY; PRT; 592 AA.

AC Q8GX14; DT 01-MAR-2003 (T-EMBLrel. 23, Created)

DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)

DE Putative glutamine amidotransferase/cyclase.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosid II; Brassicales; Brassicaceae; Arabidopsais.

NCBI_TaxID=3702;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,

RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,

RA Hayashizaki Y., Shinozaki K;

RT "Arabidopsis thaliana full-length cDNA."

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK118498; BAC43102.1; -

KW Transferase; Glutamine amidotransferase.

SQ SEQUENCE 592 AA; 64162 MW; FDD7F2B2F70CD515 CRC64;

Query Match 78.9%; Score 2270.5; DB 10; Length 592;
Best Local Similarity 80.5%; Pred. No. 1.7e-146;
Matches 434; Conservative 56; Mismatches 46; Indels 3; Gaps 2;

Qy	21	AGAGGSVVTLDDYGAGNVRVNRNARTIGFDIKDVQKPEDIIINAKELIPPGVGAFAFAPAM 80	
Db	56	ASSTSDSVTLDDYGAGNVRVNRNARTIGFDIKDVQKPEDIIINAKELIPPGVGAFAFAPAM 115	
Qy	81	DVILRLGLEALCTYIQNDRPPFLGICLGLQLLFSSESENGPIQGLGILPGRVGRFESSNG 140	
Db	116	DVILRLGLEALCTYIQNDRPPFLGICLGLQLLFSSESENGPIQGLGILPGRVGRFESSNG 175	
Qy	141	LRYPHIGWALDIKEGSAILDDVGNQHVYFVHSYRA-NAEDNKWEISSTCSYGDFFIASI 199	

Db 176 IRVPHIGWALQVKDSEILDDVGNRRHVYFVHSYRAIPSDENKDWISSTCNYGESFISII 235
 Qy 200 QKGNVHAVQHPKSGVGLSILRRFLNADSFNNKQKPMNGKASKLAKVIACLDVRAN 259
 Db 236 RRGNVHAVQHPKSGVGLSVLRRLFHPKL--PATQKMEGKASKLAKVIACLDVRN 293
 Qy 260 DNGDLVVTGQDQDVDRTERTEENVRNLGKVELAGQYVLDGADVEVFNLTGFRDPLGD 319
 Db 294 DKGDLVVTGQDQDVDRTERTEENVRNLGKVELAGQYVLDGADVEVFNLTGFRDPLGD 353
 Qy 320 LPMQLVLRASENVFVPLTVGGGIRDPDANGRYYSLEVASVYFRSGADKVSIGSDAVY 379
 Db 354 LPMQLVLRQTSKNVFLVTVGGGIRDPDANGRYYSLEVASVYFRSGADKVSIGSDAVS 413
 Qy 380 TAEYIKTVKTKGSSIEQISTVYGNQAVVSDPRVYLRKDEVEFKAIKVSHPGNG 439
 Db 414 AAEFIKSGVTKGSSLEQISRVYGNQAVVSDPRVYVNHDPDPYKVRVYVTPGNG 473
 Qy 440 EEWAVYCTVNGGREGPIGAYELAKAVEELGAGEILLNCIDCGQKGFDIDLKILSD 499
 Db 474 EEWAVYCTVNGGREGPIGAYELAKAVEELGAGEILLNCIDCGQKGFDIDLKILSD 533
 Qy 500 AVNIPVIASSGAGVADHFSVFNETNASALAAGIFHRKEVPIKAVKEHLKSGIEVRL 558
 Db 534 SVGIPVIASSGAGTPDHFSEVFETNASALAAGIFHRKEVPIQSVKEHLQEEIEVRI 592

RESULT 2
 Q9HFV7 ID Q9HFV7 PRELIMINARY; PRT; 627 AA.
 AC Q9HFV7;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Imidazole glycerol phosphate synthase.
 GN HIS7.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=5476;
 [1]
 RN SEQUENCE FROM N.A.
 RA Day T.W., Davission V.J.;
 RT "Cloning and Characterization of the CaHIS6 and CaHIS7 Genes from the
 Fungal Pathogen Candida albicans.";
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF290177; AAGL7872.1; -;
 DR HSP; Q9X0C6; 1THF.
 DR InterPro; IPR000991; GATase_1.
 DR InterPro; IPR004651; HisF.
 DR InterPro; IPR006062; His_biosynth.
 DR Pfam; PF00117; GATase; 1.
 DR Pfam; PF00977; His_biosynth; 1.
 DR TIGRFAMs; TIGR00735; hisF; 1.
 DR PROSITE; PS00442; GATASE TYPE I; 1.
 SQ SEQUENCE 627 AA; 69605 MW; 75A51D5D31A6A34C CRC64;

Query Match 41.4%; Score 1192.5; DB 3; Length 627;
 Best Local Similarity 43.6%; Pred. No. 7e-73;
 Matches 272; Conservative 104; Mismatches 147; Indels 101; Gaps 19;
 Qy 29 VTLDDYAGNVRSVRNRI--GFDIKDVKQKPEDILN---AKRLIFPGVGAFAPAMDV 82
 Db 5 IHIIDVSGNLSLSNAIKRIDSNIYIKFTNEQDFIDNDSQIEKLIFFPGVGFHFVKQ 64
 Qy 83 LIRKGLAEALCTYQNDPRFLICGLQLLFESSENGPIQGLGLIPG-----130
 Db 65 LNERKLNHLKSYIKQDRPLWGLCVGIQSIFHESESPNKGKGLGLLDNDNDDEKLL 124
 Qy 131 RVGRFESSN-----GLR--VPHIGWAL--DIKEGSAILDVGNQ-----HV 168
 Db 125 KLYKFDNDDEKFKIRGIKGVPHIGWNNIHI-----IIN--GKQTTKSLYGLNKIDKY 177

Qy 169 YFVHSTRANAEDN-----KEWISSTCSYGD--FIASIOKGNHAVQHPHE 212
 Db 178 YFVHSTRANAEDN-----KEWISSTCSYGD--FIASIOKGNHAVQHPHE 237
 Qy 213 KSGGVLGSLIRPLN-----ADSFNNKQKPMNGKASKLAKVIACL 254
 Db 238 KSGIVGLKIKIKNFLNGEKNPPTPTPTPNNLAQGVSDSSSTLDVETTLTGLTRRIIACL 297
 Qy 255 DYRANDNGDLVVTGQDQVDRER-----TEENVRNLGKVELAGQYVLDGAD 303
 Db 298 DVRTNDGDLVVTGQDQVDRER-----TEENVRNLGKVELAGQYVLDGAD 357
 Qy 304 VSFNLTGFRDPLGDPLMLQVLRASENVFVPLTVGGGIRDPD--ANGRYYSLEVAS 362
 Db 358 IIFLNITSFRNSPLKDLPLMLQVLRASENVFVPLTVGGGIRDPD--ANGRYYSLEVAS 417
 Qy 363 YFRSGADKVSIGSDAVYTAEE--YIKTVKTKGSSIEQISTVYGNQAVVSDPRVYLRK 421
 Db 418 YFQSGADKVSIGSDAVYTAEE--YIKTVKTKGSSIEQISTVYGNQAVVSDPRVYLRK 477
 Qy 422 P--DEVEFKAIKVSHP---GPNGEYAWYQCTVNGGREGPIGAYELAKAVEELGAGEIL 476
 Db 478 PTTETTTMTQIKITDPSQYGPNGEQYCYQVTSQGRKIHGALGALCLACEKLGAGEIL 537
 Qy 477 LNCIDCGQKGFDIDLKILSDAVNIPVIASSGAGVADHFSVFN--ETNASALAAGIF 535
 Db 538 LNSIDHSGSKGNFLQLLOQIKSVISPIVSIASSGAGNPQHQQVFMDCGIDAALGAGLF 597
 Qy 536 HRKEVPIKAVKEHLKSG--IEVRL 558
 Db 598 HRGEYTVNQVKYLOQEAQKMDVRL 621

RESULT 3
 Q8FNZ9 ID Q8FNZ9 PRELIMINARY; PRT; 258 AA.
 AC Q8FNZ9;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Cyclase HisF.
 GN HISF OR CE1994.
 OS Corynebacterium efficiens.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
 OX NCBI_TaxID=152794;
 [1]
 RN SEQUENCE FROM N.A.
 RA STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
 RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
 RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
 RA Usuda Y., Sugimoto S.;
 RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP005220; BAC18804.1; -;
 KW Complete proteome.
 SQ SEQUENCE 258 AA; 27106 MW; 7E57F451BFBDF8A CRC64;

Query Match 16.8%; Score 484.5; DB 16; Length 258;
 Best Local Similarity 38.8%; Pred. No. 3.7e-25;
 Matches 121; Conservative 51; Mismatches 83; Indels 57; Gaps 9;
 Qy 246 LAKRVITACLDVRANDNGDLVVTGQDQVDRTERTEENVRNLGKVELAGQYVLDGAD 305
 Db 3 VAIRVIPCLDV---DNGRVV--KGVNFE-----NLKRDAGDPVELAKRYGEGADELT 49
 Qy 306 FLNITGFRDPLGDPLMLQVLRASENVFVPLTVGGGIRDPD--ANGRYYSLEVASVYFR 365
 Db 50 FLDVSASKD---GRGTMLDVRRTAQIFLPLTVGGV-----SVEDVDQLLR 95
 Qy 366 SGADKVSIGSDAVYTAEEYIKTVKTKGSSIEQISTVYGNQAVVSDPRVYLRKPD 425
 Db 96 AGADKVSNTSATARPE-----LLSELSQRFQAQCVLSVDARRVPAGEA---140

Qy 426 EFKAIVSHFGPNGEYVAMVQCTVNGRGPRIGAYELAKAVBELGAGEILLNCIDCDGQ 485
 Db 141 -----FQPSG-----FEVTHGTGTRAGLDAVEWATGKGVGEILLNSMDGDT 186
 Qy 486 GKGFIDILKILSDAVNIPVIASSGAGVADHPSFVFNETNASALAAGIFHRKEVPKAV 545
 Db 187 KNGFDLELLKVRRAVSIPIVSIAGSGAGTAHEHFPFPAV-RAGANAVLAATIFHFGEVITTEV 245
 Qy 546 KEHLKKEGIEVR 557
 Db 246 KDAIEKAGFEVR 257

RESULT 4

Q8GDQ6 PRELIMINARY; PRT; 261 AA.
 AC Q8GDQ6; 2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Hlep (Fragment)
 OS Helicobacillus mobilis.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;
 OC Helicobacillus.
 OX NCBI_TaxID=28064;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22337798; PubMed=12446909;
 RA Raymond J., Zhaxybayeva O., Gogarten J.P., Gerdes S.Y.,
 RA Blankenship R.E.;
 RT "Whole-genome analysis of photosynthetic prokaryotes."
 RL Science 298:1616-1620(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Liollos K.G., Chu L., Ostrovskaya O., Mendybaeva N., Koukharenko V.,
 RA Gerdes S., Kyrpides N., Overbeek R.;
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AV142920; AAN87524.1; --
 FT NON TER 261
 SQ SEQUENCE 261 AA; 28552 MW; 69B03C9560582690 CRC64;

Query Match 16.5%; Score 476; DB 2; Length 261;
 Best Local Similarity 38.8%; Pred. No. 1.4e-24;
 Matches 121; Conservative 41; Mismatches 94; Indels 56; Gaps 7;

Qy 246 LAKRVIACLDVRANDGDLVVTGKQDYDVRTEENEVRNLGKPVLAGOYVLDGAEVDS 305
 Db 2 LAKRIIPCLDVHGR-----VWKG-----TNFVNLRDAGDPVELAAVYDKEGADLV 48
 Qy 306 FLNITGPRDPLGLPMLQVLRASENVFPLTVGGGIRDFTDANGRYYSLEVASFYR 365
 Db 49 FLDITASSD---GRAIMLDVRRTAEBVFIPFTVGGGLR-----TVEDIREMLK 94
 Qy 366 SGADKVSIGSDAVYTAEEYIKTVGKTSIEQISTVYGNQAVVVSIDPRRVYLKRPDEV 425
 Db 95 AGADKISLNTSAVQTPQ-----LIGDSAWKFGSQCVVAIDARR---RRDEE- 138
 Qy 426 EFKAIVSHFGPNGEYVAMVQCTVNGRGPRIGAYELAKAVBELGAGEILLNCIDCDGQ 485
 Db 139 -----GRLEGHEVYTHGGRKPTGIDVLEWARKVBEKGCEILLTSMDDGDT 185
 Qy 486 GKGFIDILKILSDAVNIPVIASSGAGVADHPSFVFNETNASALAAGIFHRKEVPKAV 545
 Db 186 KDGFDIPLTRAVSAVKIPVIASGGVGNLEHICEGLTAGKADALAASIFHYKEYTIRET 245
 Qy 546 KEHLKKEGIEVR 557
 Db 246 KEYLRGKGHVHR 257

RESULT 5

Q8G6F7 PRELIMINARY; PRT; 256 AA.
 AC Q8G6F7; 2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Hlep cyclase.
 GN HISP OR BL0885.
 OS Bifidobacterium longum.
 OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
 OC Bifidobacteriaceae; Bifidobacterium.
 OX NCBI_TaxID=216816;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=NCC 2705;
 RX MEDLINE=22294977; PubMed=12381787;
 RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Beiger B.,
 RA Pessi G., Zwhilen M.-C., Desiere F., Bork P., Delley M.,
 RA Fridmore R.D., Arigoni F.;
 RT "The genome sequence of Bifidobacterium longum reflects its adaptation
 to the human gastrointestinal tract."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
 DR EMBL: AE014691; AAN24505.1; --
 KW Complete proteome.
 SQ SEQUENCE 256 AA; 27551 MW; 7CE149FCB0759B5C CRC64;

Query Match 15.3%; Score 439; DB 16; Length 256;
 Best Local Similarity 35.0%; Pred. No. 4.6e-22;
 Matches 111; Conservative 46; Mismatches 92; Indels 68; Gaps 9;

Qy 246 LAKRVIACLDVRANDGDLVVTGKQDYDVRTEENEVRNLGKPVLAGOYVLDGAEVDS 305
 Db 3 LAVRVIPCLDVRANDGDLVVTGKQDYDVRTEENEVRNLGKPVLAGOYVLDGAEVDS 49
 Qy 306 FLNITGPRDPLGLPMLQVLRASENVFPLTVGGGIRDFTDANGRYYSLEVAS 362
 Db 50 FLDVITASSHRN-----TMIDVVSRTABQVFIPTVGGVTRPEDVD-----S 92
 Qy 363 YFRSGADKVSIGSDAVYTAEEYIKTVGKTSIEQISTVYGNQAVVVSIDPRRVYLKRP 422
 Db 93 LLRCGADKGVNTAAI-----NDPSLISRVADRFGQVQLVLSVDARR----- 134
 Qy 423 DEVEFKAIKSHFGPNGEY--AWYQCTVNGRGPRIGAYELAKAVBELGAGEILLNCI 480
 Db 135 -----EKGEQTSQSGFEVTTMGGRKSTGIDAIWVVKRAEQIAGAGEILLNSM 180
 Qy 481 CDCQCKGKGFIDILKILSDAVNIPVIASSGAGVADHPSFVFNETNASALAAGIFHRKEV 540
 Db 181 DAGTKEGDFLEMRIRAVRKEVKIPIIASGAGKVEDFPFPAI-EAGADAVLAASVFFHYGIL 239
 Qy 541 PIKAVKEHLKKEGIEVR 557
 Db 240 TIADVKAEKKGHYTVR 256

RESULT 6

Q8R883 PRELIMINARY; PRT; 203 AA.
 AC Q8R883; 2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Glutamine amidotransferase.
 GN HISH OR TTE2135.
 OS Thermoanaerobacter tengcongensis.
 OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
 OC Thermoanaerobacteriaceae; Thermoanaerobacter.
 OX NCBI_TaxID=119072;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=MB4 / JCM 11007;
 RX MEDLINE=21992816; PubMed=11997336;
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.,
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439 (2002).
DR ENBL; AE014962; AAN58946.1; -.
KW Complete proteome.
SQ SEQUENCE 251 AA; 26874 MW; 1C27D132D16BFBEA CRC64;

Query Match 14.6%; Score 421; DB 16; Length 251;
Best Local Similarity 34.8%; Pred. No. 7.5e-21;
Matches 110; Conservative 47; Mismatches 89; Indels 70; Gaps 10;

QY 246 LAKRVIACLDVRANDGLVVTGDDQYDVRTEENEVRNLGKPVLAGOYLDGADSV 305
Db LKRIIPCLDVK-----DGRVKGVF-----VNLTDVGPVDAARAYYAGCDELV 48
QY 306 FLNITGPRDFPLGLDPLQVLRASENVFPLTVGGGIRDTDANGRIYSSLEVASFYR 365
Db FLDTATSD---NRETTVDVRHVADQVFPTVGGGIRSVDDMN-----KMLK 94
QY 366 SGADKVSIGSDAVYTAEYIKTGVTGKSSIEQISTVYGNQAVVSDPRRVYLRKP 425
Db AGADKAVNSSAI-----ANPLIKDCAEKFGSCVVAIDAR---KEADD- 137
QY 426 EPFAIKVSHPGNGEYAVYQCTVNGGREGPIGAYELAKAVELGAGETLLNCIDCQ 485
Db 138 -----SWH-VYVAGRKDTGIDLLANVKEAVOLGAGETLLTSMKDG 179
QY 486 KGKGFIDIDLIKLVISDAVNPVIASSGAGVADHFSEVFNENASAAAGIFHRKEVPI 542
Db 180 KSGFDLMLNVAQLADIPVIASSGAGNMEHVEIFKTPATGALAAASIFHYGEVSIADT 239
QY 543 -KAVKEHLKKEGIEVR 557
Db 240 KKAMKEH-----GIEVR 251

RESULT 10
Q8DJN7
ID Q8DJN7 PRELIMINARY; PRT; 281 AA.
AC Q8DJN7
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Inidazoleglycerol-phosphate synthase, cyclase subunit.
GN HSF OR TLR1185.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1";
RL DNA Res. 9:123-130 (2002).
DR ENBL; AP005373; BAC08737.1; -.
KW Complete proteome.
SQ SEQUENCE 281 AA; 30011 MW; A8AB0E9F7430372 CRC64;

Query Match 14.3%; Score 412.5; DB 16; Length 281;
Best Local Similarity 34.3%; Pred. No. 3.4e-20;
Matches 108; Conservative 47; Mismatches 97; Indels 63; Gaps 7;

QY 246 LAKRVIACLDVRANDGLVVTGDDQYDVRTEENEVRNLGKPVLAGOYLDGADSV 305
Db 3 LAKRIIPCLDVKAGR-----VKGKVF-----VNLRDAGDPVLAQYNAAGADELV 49
QY 306 FLNITGPRDFPLGLDPLQVLRASENVFPLTVGGGIRDTDANGRIYSSLEVAS 362
Db 50 FLDTATHEERNI-----LIDVYTAQVFIPLTVGGGIQ-----SLTMKD 92
QY 363 YFRSGADKVSIGSDAVYTAEYIKTGVTGKSSIEQISTVYGNQAVVSDPRRVYLRKP 422
Db 93 LLRAGADKVSLSAAVRQPD-----LVNQASDRFGAQCIIVVADARREPCAP 140
QY 423 DEVEFPAIKVSHPGNGEYAVYQCTVNGGREGPIGAYELAKAVELGAGETLLNCIDC 482
Db 141 DQPR-----WQVYVGRGREATGLDAVAVEMAKRGAGELLVTSM 182
QY 483 DGKGFIDIDLIKLVISDAVNPVIASSGAGVADHFSEVFNENASAAAGIFHRKEVPI 542
Db 183 DGTQAGYDLELFAERVEIPVIASSGAGTCEHRAALVEGKAEALLASLHYGOLTI 242
QY 543 KAVKEHLKKEGIEVR 557
Db 243 AQIKGYLHQHQPVR 257

RESULT 11
Q8KF56
ID Q8KF56 PRELIMINARY; PRT; 201 AA.
AC Q8KF56
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Amidotransferase Hish.
GN HSH OR CT0476.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OX Chlorobium.
OX NCBI_TaxID=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TLS / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; PubMed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
RA Vamathavan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT "The complete genome sequence of Chlorobium tepidum TLS, a
RT photosynthetic, anaerobic, green-sulfur bacterium.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514 (2002).
DR ENBL; AS012824; AAM71718.1; -.
DR TIGR; CT0476; -.
DR InterPro; IPR000991; GATase_1.
DR Pfam; PF00117; GATase; 1.
DR PROSITE; PS00442; GATASE TYPE I; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 201 AA; 22259 MW; D1873B5D78F6BF6B CRC64;

Query Match 14.1%; Score 407; DB 16; Length 201;
Best Local Similarity 43.0%; Pred. No. 4.8e-20;
Matches 86; Conservative 38; Mismatches 72; Indels 4; Gaps 4;

QY 28 VVTLLDYGAGNVRVSNARTLGFIDIKVQKQEDILNAKRLIFPGYGAFAPANDVLRKG 87
Db 1 MVFIADYGAGNVRVSNARTLGFIDIKVQKQEDILNAKRLIFPGYGAFAPANDVLRKG 60
QY 88 LAEALCTYIQNDPPFLGICLGLQLLPESSEBNGPIQGLGLIPGRVGRFESSNGLRVP 147

DB 61 FDEAIREHIDKRSVLGICLGMQLFLSESEMGAYKGLDIVPKVLRFTSSD-KIPQIG 119
QY 148 WHALDIKEGSAIIDDVGNQ-HVTVHSYRANADNKWISSTCSY-GDDPIASIQGNVH 205
DB 120 WNSVDYCKDSVLFNRPDQSYFYFVHSYYC-APDEPESVAATFFAGKFKFCSAIEKNGIF 178
QY 206 AVQFHPKSGGVGLSLRRF 225
DB 179 AVQFHPKSGSGLQVLKNF 198

RESULT 12

Q8FY07 PRELIMINARY; PRT; 261 AA.
AC Q8FY07;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Inidazoleglycerol phosphate synthase, cyclase subunit.
GN HSF OR BR2085.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1330 / Biovar 1;
RX MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tettelin L.H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Helling S.M., Boyle S.M., Fraser C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts."
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153 (2002).
DR EMBL; AE014495; AAN30975.1; -
DR TIGR; BR2085; -
KW Complete proteome.
SQ SEQUENCE 261 AA; 27474 MW; DECA75E3993D1720 CRC64;

Query Match 14.1%; Score 405; DB 16; Length 261;
Best Local Similarity 35.1%; Pred. No. 9.9e-20;
Matches 110; Conservative 41; Mismatches 104; Indels 58; Gaps 7;

QY 246 LAKRVIACLDVRANDGDLVVTGKDQYDVRTEENVRNLGKPVLAGQYILDGDEV 305
DB 3 LKARVIFCLDVK-----DGRVVGKGVNF-----VDLIDAGDPVEAARAYDAAGDEL 49
QY 306 FLNITGFRDPLGLPLQLQVLRASENVFPLTVGGGIRDFDTANGRYYSLSLEVSEYFR 365
DB 50 FLIDTASSD---NRETIFDVVARTAEQCFLPLTVGGGVQVADIR-----KLL 95
QY 366 SGADKVISGDAVYTAIEYKTVTKGKSIEQISTVYGQAVVSTDPVRVLRKDEV 425
DB 96 AGADKVSINTAAVKNP-----FVAEADKFGNQCIIVAID----- 131
QY 426 EFKAIKSHPGNCEEYAWYQCTVNGREGRPICAYELAKAVELGAGEILLNCIDCGQ 485
DB 132 ---AKKVSAGENDR-----WEIFTHGGROPTGIDAVEFAKQVDLAGEILLTSMRDGT 184
QY 486 KGKFDILIKLISDAVNPVIASSGAGVADHFSVFNETNASALAAAGIFHRKEVPKAV 545
DB 185 KAGYDVALTRAVDSVRAPVIASGGVGTLDHLVAGIRGDHATAVLAASIFHFGTYTIGEA 244
QY 546 KEHLKKEGIEVRL 558
DB 245 KRYMAEAGIPMLR 257

RESULT 13

Q8ESS0 PRELIMINARY; PRT; 201 AA.
AC Q8ESS0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Amidotransferase (EC 2.4.2.-).
GN OB0549.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments."
RL Nucleic Acids Res. 30:3927-3935 (2002).
DR EMBL; AP04594; BAC12505.1; -
KW Glycosyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 201 AA; 22321 MW; 36B8F4B049ECAAE CRC64;
Query Match 13.8%; Score 396; DB 16; Length 201;
Best Local Similarity 40.2%; Pred. No. 2.7e-19;
Matches 80; Conservative 44; Mismatches 71; Indels 4; Gaps 3;

QY 28 VVTLLDYGNVRSVNRNARTLGFIDKQVQEDILNARLPPGPGVGAAPAMDLVRKG 87
DB 1 MTAIIDYAGNIKSLQFALDKLNKHSIVTTEAEIKQADSIIILPGVGAFKDAEAIKRLQ 60
QY 88 LAELCTYQNDRPFLGICGLQLLPFESSENGPIQGLIGRVRGFSNGLRVPHG 147
DB 61 LOSVIOEEAKQKPLIGICGLQGLFQESLENGDWGLLKGSIKRI--SGEVKVPHG 118
QY 148 WHALDIKEGSAIIDD-VGNQHVYFVHSYRANADNKWISSTCSYGDFFIATQKGNVHA 206
DB 119 WNTLDIQQASPLFDSKLENPPYVYFVHSYAVSPEENTLLASS-QYGLIPAIQKGNITG 177
QY 207 VQFHPKSGGVGLSLRRF 225
DB 178 MQFHPKSGFGIELKRY 196

RESULT 14

Q939J6 PRELIMINARY; PRT; 201 AA.
AC Q939J6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical 22.7 kDa protein.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=81-176;
RX PubMed=11461915;
RA Thibault P., Logan S.M., Kelly J.F., Brisson J.R., Ewing C.P.,
RA Trust T.J., Guerry P.;
RT "Identification of the Carbohydrate Moieties and Glycosylation Motifs
RT in Campylobacter jejuni Flagellin."
RL J. Biol. Chem. 276:34862-34870 (2001).
DR EMBL; AY034084; AAK58487.1; -
DR InterPro; IPR000991; GATase_1.
DR Pfam; PF00117; GATase_1.
DR PROSITE; PS00442; GATASE_TYPE_I; 1.
KW Hypothetical protein.
SQ SEQUENCE 201 AA; 22732 MW; BC60525EDABC2A3C CRC64;

Query Match 13.7%; Score 394; DB 2; Length 201;
Best Local Similarity 40.5%; Pred. No. 3.7e-19;
Matches 81; Conservative 35; Mismatches 80; Indels 4; Gaps 3;

QY 28 VTLLDYGAGNVRVSRNAITLGFIDKDVQKPEDILNAKELIPFGVGAFAPAMDLIRKGL 87
DB 1 MIALIDYKAGNLNSVAKAFKIGAINPIAKNPDKLQKADKLLPFGVGSPKEMKNLKELG 60
QY 88 LABALCTYI-QNDRPFLGICLGLQLLPESSEENGPIQGLGLIPRVRGRPESSNGLRVPHI 146
DB 61 FIEALKEQVLVQKKPILGICLGLQLPFLRGYEGVCEGLGFIEGVVVKFEDLNLIKPHM 120
QY 147 GWHALDIKEGSAILDVGNQ-HVYFVHSYRANAEDNKWISSTCSYGGDDFIASIQKGNVH 205
DB 121 GWNELEILKQDPLYQGINKNSDFYVHSFYVCKKD--EFVSAKAQYGHKEVASLQKDRIP 178
QY 206 AVQFHPKESGGVGLSILRRP 225
DB 179 ATQFHPKESQNLGLKLENF 198

RESULT 15

Q8TS91 PRELIMINARY; PRT; 202 AA.
AC Q8TS91;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Imidazoleglycerol-phosphate synthase, subunit H.
GN HISH OR MA0913.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR ENBL; AE010756; AAM04346.1; -
DR InterPro; IPR000991; GATase_1.
DR Pfam; PF00117; GATase; 1.
DR PROSITE; PS00442; GATASE_TYPE_I; 1.
KW Complete proteome.
SQ SEQUENCE 202 AA; 21860 MW; 122A2875EBFC371C CRC64;

Query Match 13.7%; Score 393; DB 17; Length 202;
Best Local Similarity 44.3%; Pred. No. 4.4e-19;
Matches 89; Conservative 30; Mismatches 74; Indels 8; Gaps 5;

QY 29 VTLLDYGAGNVRVSRNAITLGFIDKDVQKPEDILNAKRIIPFGVGAFAPAMDLIRKGL 88
DB 4 IVIDYGLNLSVQKGLHVGANPAISGNPEILTADGILPFGVGAFIDAMKCLI--PL 61
QY 89 AEALCTVIQNDRPPLGICLGLQLLPESSEENGPIQGLGLIPRVRGRPESSNGLRVPHIGW 148
DB 62 KGVAIEFAESGKPMGLICLGLQVLMSSSEGRLTGGLDLIQGRVLRPKSE-LKVPFMGW 120
QY 149 HALDIKEGSAILDVGN-QHVYFVHSYRANAEDNKWISSTCSYGGDDFIASI--QKGNVH 205

Db 121 NNIRIKQDHPHLPKGISDGSFVYFVHSY--YVDTTAENTLASCEYGLDFASVSVNSKGNVM 178
QY 206 AVQFHPKESGGVGLSILRRFL 226
DB 179 GTQFHPKESGTTGLKILKNFV 199

Search completed: January 20, 2004, 16:27:33
Job time : 43 secs

